Production and Perception
of (Non-)Seriousness in German Conversation

By

Brandy J. Trygstad

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(German)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: August 23, 2018

Dissertation is approved by the following members of the Final Oral Committee:

Joseph Salmons, Professor, Language Sciences
Andrea Golato, Dean, The Graduate College, Professor, Department of Modern Languages, Texas State University
Robert Howell, Professor Emeritus, Department of German, Nordic, and Slavic
Eric Raimy, Professor, Department of English
Manon van de Water, Professor, Department of German, Nordic, and Slavic

For Ellis M. Taft, who would have loved every word.

Fair winds and following seas.

Acknowledgements

First and foremost, I wish to thank the anonymous participants who volunteered their time and effort and, in some cases, who opened their homes to me for recording. This work would not exist without your willingness to offer me the use of your conversational and survey data in the name of science. Thank you, from the bottom of my heart, for your trust in me and your immense contributions to the study of the German language.

I would like express deepest gratitude to my advisor, Prof. Joe Salmons. Without his careful guidance and clear insight, without his willingness to step in to pull me out of limbo and get me back on track, this research would not have been possible. He was a constant source of inspiration and motivation, always challenging me to do the best I could, and then do better. His patience and unwavering support at all stages of this project make him the best *Doktorvater* one could ever dare to dream of. Thank you, Joe, for everything.

I also wish to thank my committee members, Prof. Andrea Golato, Prof. Robert Howell, Prof. Eric Raimy, and Prof. Manon van de Water, for their support and insightful feedback. In particular, thank you to Andrea, for introducing me to CA and always pushing me to sharpen and improve my analyses, and to Eric, for inspiring and encouraging me to take a more precise approach to the acoustic features of conversation. A very special thanks as well to Prof. Tom Purnell, whose guidance in research design, analysis, and presentation proved invaluable.

Thank you to Prof. Margret Selting for hosting me at the Universität Potsdam in 2012-13, and to the Fulbright Program, for funding the data collection phase of my research during that time. I also thank my colleagues at the Universität Potsdam, Maxi Kupetz, Uwe Küttner, and Jana Scheerer, as well as Prof. Selting, for welcoming me without reservation into the research group, and for their feedback on the early transcription and analysis of the data used in this study.

Thank you also to: Prof. Celia Kitzinger, who first drew my attention to and urged me to consider jokes and non-seriousness in German; Prof. Jürgen Trouvain, who provided key insights which vastly improved the design of my survey; Prof. Dagmar Barth-Weingarten and Prof. Margret Selting, who offered valuable feedback on my understanding of prosody in interaction and asked challenging yet inspiring questions; Prof. Salvatore Attardo, who kindly provided materials to supplement my understanding of his work; Prof. Thorsten Huth, who gave incredibly helpful and insightful commentary on my earliest analyses of non-seriousness; Prof. Mark Louden, who saw me through the early stages of this work; Prof. Manon van de Water and Dr. Jeanne Schueller, who supported me when I needed it most; Prof. J. P. Spicer-Escalante and Prof. Maria Luisa Spicer-Escalante, who have been my dearest mentors throughout my academic career; Prof. Nancy Bell, who introduced me to the joy of linguistics; and all the many scholars and colleagues who offered their feedback and insight over the years, whether at conferences, in collaboration, or over coffee. Though I may have missed your name here, your contributions have been neither forgotten nor taken for granted.

To my friends, who not only put up with me over the last century, but who also offered their assistance at key points throughout, thank you so much for all you have done to help bring this project to fruition. There are more of you than I can possibly list, but I would like to thank in

particular: Mike Olson, Beth Tucker Long, and Nicholas Williams, for your utterly indispensable assistance with the technical side of this project, without which I simply could not have achieved the required levels of precision in analysis; Jae DiBello Takeuchi, for spending hours upon hours with me in the library and giving extensive feedback on my writing; Maya Nigrosh, for your support and encouragement, not only in this endeavor, but also in all our zany projects, from cement fists to wax wings, and for putting up with the fluff and glitter; Marie Stolzenburg, for proofreading and providing refreshments; Daniel Pannicke, for the most well-timed postcard in all of history; and Matthias Höll, who is the only one who will get this joke.

A special thank you goes to Jenn Schicker, who more than once dropped everything to come to my side when I needed her most. Thank you not only for giving invaluable feedback and discussing ad nauseam the thorniest problems in my data, but also for taking care of me, physically and emotionally. Though the miles separate us, I know you are never truly far away.

To my family, thank you for your support and encouragement, not only during these last few years, but more importantly, from the earliest days of my education. Thank you especially to Grandpa, who taught me that there's no sense in starting anything if you don't intend to do your very best, and that no job is finished until the paperwork is done. Thank you also to my parents, who got me through school and sent me off into the world. To Mom, Dad, and Sonya, I am so happy to tell you that my "paper" is finally done. To my chosen family, Ruthie, Tinkerbell, and Jace, thank you for laying on my laptop keyboard whenever I forgot that it was bedtime.

There are so many people who have contributed to this work, in ways big and small. Much like a child, it truly takes a village to raise a dissertation. Without doubt, I have missed somebody here, but that does not at all diminish your contributions. I am truly, deeply appreciative of the support of all those, named and unnamed, who in any way helped me along this journey.

Finally, last but in no way least, I wish to offer my most heartfelt gratitude to Adam Bauknecht. It is impossible to put into words how much I appreciate your support and encouragement during these last few years. There are some gifts that can never be repaid; your unwavering patience and unshakeable belief in me is one such gift. Thank you, Adam, for everything. Thank you.

Table of Contents

Production and Perception of (Non-)Seriousness in German Conversation

Tables and Figures

Data Excerpts

Abstract

Previous research using Conversation Analysis (CA) has documented the sequential features of non-seriousness (e.g., humor and jokes) in naturally occurring conversation. On the other hand, studies utilizing sociophonetic methods have analyzed data from a variety of sources, yet rarely considered non-seriousness in conversational data. There is a need to more precisely define the unique acoustic features of non-serious utterances in everyday conversations, and how these might come to bear on the ability of one's interlocutor to accurately identify a given utterance as non-serious in order to be able to respond differentially and appropriately.

To advance our understanding of non-seriousness, this study pursues two main types of inquiry, namely, a production analysis and a perception experiment. The production arm of the study makes use of both CA and sociophonetic methodologies. Each speech excerpt was first analyzed sequentially using CA to verify its status as a non-serious utterance, and to locate a serious counterpart from the same speaker in contiguous speech immediately preceding the non-serious utterance. The excerpts were then analyzed acoustically to measure changes in speakers' non-serious utterances from baseline serious talk. These same excerpts were later used in a perceptual experiment, allowing comparison of the acoustic analyses with responses on a survey in which participants were asked to differentiate between serious and non-serious utterances.

The production results suggest that speakers' non-serious utterances differ acoustically from baseline serious utterances, while the results of the perceptual survey suggest that these changes are salient to participants listening to recorded conversations. In contrast to previous studies which have attempted to locate and define a set of acoustic features manipulated in a consistent way to mark non-seriousness, I show that speakers may manipulate a constellation of acoustic features in a variety of ways which may vary significantly by speaker. In other words, there may not be a singular marker of non-seriousness for all speakers. The results of this study allow us to approach a fuller understanding of the interplay between sequence and prosody, a result accessible only through the proposed combined approach harnessing the best practices of both CA and sociophonetics.

Chapter 1

Responding to Non-Seriousness: Preliminary Observations

> "It is not an excusable error to hear an assertion as 'serious' and use the device appropriate if the assertion were heard as 'joke,' and vice versa." (Sacks, 1972, p. 44)

1.1. Introduction

In early work that inspired this study, I noted that participants frequently seemed to take on a different "voice" when making utterances which were identifiable on the basis of content, context, and how they were treated by co-conversationalists as jokes or instances of humor. Because this alternate voice could not be said to differ from the surrounding talk on any one variable (such as special stress, pitch peaks, or changes in intensity), a recurrent problem with this observation was how to transcribe the audible phenomenon using the Jeffersonian transcription conventions of Conversation Analysis (CA). This problem was ultimately compounded by my realization that the key elements of what I came to call "smile voice" seemed to consist of a bundle of acoustic variables, which in turn could not be precisely parsed and identified.

However, which acoustic features those might be and how they were being manipulated remained open questions, so the transcription problem remained. Lacking clarity on the nature of this acoustic phenomenon, the latter problem was initially addressed with a kludge: I simply began using the symbol ☺ in my transcripts to mark utterances which a) were made with vocal

qualities which I perceived as appreciably different from the surrounding talk, albeit without insertion of laugh particles into the utterance itself (cf. Hepburn & Varney, 2013; Jefferson, 1985), and b) were treated as humorous (e.g., with laughter subsequent to the utterance) by some or all participants in the interaction.

In time, I informally adopted the moniker "smile voice" to describe the phenomenon. Although I later learned that this same (or similar) term has been used by some researchers to describe moments when speakers sound as if they are smiling while speaking (e.g., Couper-Kuhlen, 2004; Holt, 2013b; Linneweber, 2016; cf. "speech-smile" in Kohler, 2008; "happy talk" in Tartter, 1980) , this was not my intended interpretation per se (i.e., that speakers were literally smiling). Rather, this term was used as a shorthand reference to the ☺ symbol I used in the earliest transcripts to mark instances of utterances treated as humorous which differed acoustically from surrounding talk.

As I amassed a collection of utterances marked with the ☺ symbol, I hypothesized that the phenomenon was likely to consist of a variety of features, both acoustic and nonacoustic, which speakers may manipulate in order to create the impression of something akin to "smiling with a wink and a nod" while speaking, thereby signaling to their co-conversationalists that their utterance was not intended to be treated as serious. Such marking is of obvious consequence in sensitive social interactions involving actions such as teasing. However, these features have not been identified in the extant literature, a gap which this project will begin to fill by investigating a) whether and how participants in naturally occurring interactions mark utterances as non-serious, and b) whether untrained listeners can identify "smile voice". In other words, if "smile voice" exists as a salient feature of conversation, speaker-listeners will be able to both produce it as a means of marking humor in naturally occurring interactions, and also perceive it as a cue for

humor.

1.2. Background

The study of humor in speech has tended to focus generally on two divergent areas of inquiry: the rhetorical aspects of humor, and its linguistic features. These treatments of humor are frequently based on the ideal speaker-listener, without consideration of the interactional aspects of humor as they exist outside the confines of the analyst's imagination and/or laboratory. However, more recent research has revealed some interactional features of humor, especially through use of CA. In light of this research tradition, it is clear that there is a gap in our understanding of the interplay between the interactional and phonetic features of humor or "joking", especially as found in everyday conversations.

1.2.1. Theories of Humor

Traditional theories of humor are rooted in philosophy and psychology / psychoanalysis, with an emphasis on approaching a working definition of what constitutes humorousness, as well as on differentiating between types of humor (for example, irony versus so-called "canned" jokes) and how each is constructed so as to achieve humorousness. These treatments of humor are generally based on how speakers come to appreciate the inherent humorousness of a given utterance, with minimal consideration given to the interactional aspects of humor. Veatch (1998) is a representative example of such a study, wherein he proposes a theory of humor which defines three core properties of humor, namely apparent absurdity, violation of norms, and incongruity. Similarly, Attardo and Raskin (1991) propose their framework for a model and taxonomy of humor, dubbed the General Theory of Verbal Humor. An overview of the history and development of theories of humor, from Plato to Freud, is provided in the seminal work by Attardo (1994), in addition to a description of the semantic and rhetorical components of jokes,

and the various types of humor.

Building on these general theories of humor, much research has been done on defining and differentiating between the various types of humor. For example, Attardo (2000, 2002) proposes theories of irony rooted in philosophy and rhetoric, while also defining its unique markers which differentiate it from other varieties of humor. Similarly, Kreuz (1996) analyzes the cues of irony, again grounded in rhetoric, while Haverkate (1990) focuses on the uses of irony in rhetoric and discourse. Other researchers have focused on additional varieties of humor, as in Dynel (2009), with a detailed analysis of humorous garden paths, and Sherzer (1985), who seeks to differentiate puns from jokes.

Taking a slightly different tact, Norrick (2003) applies these rhetorical theories of humor to an analysis of conversational humor. He ultimately concludes that any theory of humor must account not only for performed jokes, but also for wordplay and other varieties of conversational humor. In other words, he argues that, although the study of humor grounded in a philosophical and/or psychological approach is vital to our theoretical understanding of how humor functions, we must also consider how speakers achieve humorousness in conversation and everyday interaction, and incorporate those findings into any overarching theoretical framework. Although this study does not broach the larger question of what role, if any, the results could play in refining our theoretical understanding of humor, it nonetheless contributes a body of evidence towards that ultimate goal.

1.2.2. Linguistics of Humor

Turning now to research on the linguistics of humor, the following selection of studies offers an overview of how linguists have generally approached the study of humor. Research in this area has explored the acoustic, syntactic, and semantic features of the production of humor

using a wide variety of data types (cf. Attardo, 2017). We consider some examples of each, moving from surveys and breaching experiments, to laboratory studies, and finally, data collected from television and radio programs.

Building on the classic theories of humor described above, some linguists have sought to test the rhetorical and/or psychological theories concerning what makes a joke humorous. For example, researchers have conducted surveys designed to test variations on sarcasm and irony in order to determine what makes a given utterance identifiable as such (e.g., Gerrig & Goldvarg, 2000; Kreuz & Link, 2002; Kreuz & Roberts, 1995), as well as to gather participants' perspectives on how humor is used and why (Chiaro, 2009). Taking a slightly different approach to this question, Bell (2009) conducted a classic breaching experiment, in which she approached strangers to tell a canned joke designed to fail. This study was designed to test how interlocutors would respond to such failed attempts at humor in order to lay bare the underpinnings of humorousness itself.

Diverging from linguistic studies concerning the rhetorical features that render a joke humorous, some researchers have trained more focus on the production and perception of humor. The focus here has overwhelmingly been towards consideration of the acoustic features of laughter (e.g., Bachorowski, Smoski, & Owren, 2001; Kipper & Todt, 2003; LaGreca, Laforest, Alioua, & Vincent, 1996; Vettin & Todt, 2004), rather than on the humorous utterances proper. Yet a variety of studies using laboratory data have sought to define the various acoustic, syntactic, and semantic features of humor, in terms of both production and perception. Broadly speaking, these laboratory studies utilized data collected from recordings of participants reading aloud from a script, participating in interviews with the researcher(s), and participating in conversations with other participants.

Several studies using recorded data read aloud from a script have focused solely on acoustic production of ironic or sarcastic utterances (e.g., Anolli, Ciceri, & Infantino, 2000; Attardo & Pickering, 2011) or on perception of the same (e.g., Voyer & Techentin, 2010). All of these examples used full sentence-level utterances in their recordings, which were produced by volunteer speakers who were not professionally trained speakers. The recordings were then analyzed acoustically, with results indicating acoustic variation between types of irony (Anolli et al., 2000)) or refining our understanding of "timing" in humor (Attardo & Pickering, 2011), as well as indicating that the sarcastic tone of voice could be identified by listeners (Voyer & Techentin, 2010). In addition to these, a series of studies relying on script data investigated both production and perception of humor (e.g., Auberge & Cathiard, 2003; Cheang, 2008; Cheang & Pell, 208, 2009, 2011; Rockwell, 2000; Schaffer, 1981; Tartter, 1980; Tartter & Braun, 1994). All of these studies also used full sentence-level utterances, with the exception of Tartter (1980) and Tartter and Braun (1994), where nonce syllables were recorded; in the case of Auberge and Cathiard (2003) and Rockwell (2000), professional speakers were recorded. As with the previous studies, the results identified some acoustic features of humorous and/or sarcastic utterances, while also establishing that participants listening to the same recordings could differentiate between the serious and non-serious utterances.

Fewer studies have utilized data collected during interviews in a laboratory. For example, Kersten (2009) investigates how participants narrate humorous stories in a second language, while Pickering et al. (2009) analyze the prosodic markers of humor in canned jokes. On the other hand, researchers have invited multiple participants to participate in informal conversations amongst themselves, albeit in a laboratory setting. These studies have generally focused on the acoustic markers associated with ironic or sarcastic utterances (e.g., Attardo,

Pickering, & Baker, 2011; Archakis, Giakoumelou, Papazachariou, & Tsakona, 2010; Bryant, 2010, 2011; Nwokah, Hsu, Davies, & Fogel, 1999). Graham, Papa, and Brooks (1992) describe various functions of humor in interaction under the framework of interactional competence, while Bertrand and Priego-Valverde (2011) analyze the prosodic features of utterances treated as non-serious; the latter study utilizes CA methodology to determine which utterances to measure acoustically. All of the studies cited here focused solely on production of humor, as opposed to perception, and described variation in the acoustic features of humorous utterances as compared to surrounding serious talk.

Finally, in an attempt to capture more naturalistic data, some linguists have turned to data culled from television and radio programs. Attardo, Eisterhold, Hay, and Poggi (2003), Purandare and Litman (2006), and Urios-Aparisi and Wagner (2011) all measured the acoustic features of irony and/or sarcasm as performed on various scripted television programs (American situation comedies, *Friends*, and *Sex and the City*, respectively). In addition to analyzing the acoustic features of ironic speech, in their study utilizing radio data, Bryant and Fox Tree (2002, 2005) also tested perception of irony using a survey. In the case of both television and radio data, the researchers again documented evidence for prosodic markers of humorous utterances, while Bryant and Fox Tree (2002, 2005) additionally found evidence that these markers were salient to participants in a survey.

Taking this representative sampling in aggregate, the results of these studies overwhelmingly concur that: a) there is variation in acoustic production when making a joking or sarcastic utterance as compared to baseline, serious utterances, and b) these acoustic features are perceptible and salient for listeners. However, there is also huge variation in the precise results reported for each study, as described in detail in Chapter 3 (production) and Chapter 4

(perception). Although the detailed results taken at face value can seem at times contradictory, this is actually not surprising, given that there is also huge variation in the sources of the data, the collection methods, and the analyses.

Ultimately, however, the overarching problem is that we are not using naturally occurring conversational data for the study of non-seriousness (with the key exception of research on its sequential features; cf. Section 1.2.3.), instead relying on data which range from entirely unnatural (breaching and scripts) to idealized performances (interviews and programs on television or radio). These methods of documenting the linguistic features of production and perception of non-seriousness, relying on data collected almost exclusively in laboratory settings, leave much to be desired. Given the artificial nature of such data, with participants mimicking what they believed a genuine "smile", joking, sarcastic, and/or humorous voice should sound like, it is difficult to determine whether or to what degree a) the study participants' production accurately reproduces natural conversation, or b) the participants in the perceptual studies would have similar success in identifying joking or sarcasm in naturally occurring interactions. Indeed, we know from the CA literature that there is a vast difference in how people *believe* they speak versus how they *actually perform* in naturally occurring interactions (e.g., Golato, 2003). Therefore, despite the promising results of these studies, it is difficult to make any claims about how these results can be translated into conclusions about production and perception of humor in everyday conversation, if at all. This leaves open the question of how speakers potentially mark an utterance as humorous in naturally occurring conversations, and how conversationalists perceive that an utterance is humorous, beyond (or in the absence of) clear contextual clues.

1.2.3. Conversation and Humor

As described in the previous section, the vast majority of linguistics research on humor

has relied on data collected in laboratory settings, often under carefully controlled experimental conditions (script reading, interviews, etc.). By contrast, CA has made vital inroads in examining the interactional and sequential features of non-seriousness in naturally occurring conversation. Following the pioneering methodological approach proposed by Sacks, Schegloff, and Jefferson (1974), these studies rely nearly exclusively on naturally occurring conversational data (as explicated in more detail in Chapter 2). CA studies have traditionally hewed to three key foci, namely: laughter, sequential features of non-seriousness, and interactional uses of non-seriousness (cf. Glenn & Holt, 2017). I provide a representative survey of such studies here.

A particularly rich vein of research utilizing CA methodology has documented the key role of laughter in everyday conversation. The most extensive earlier work in this area was done by Jefferson, who, through a series of studies, examines the various functions and effects of laughter and how it is sequentially located in the course of a conversation in a series of articles (1972, 1979, 1984, 1985, 2004, 2010; Jefferson, Sacks, & Schegloff, 1987). She finds, for example, that there may be a gendered distribution of who laughs when (2004), such that women more frequently tend to laugh in response to men's laughter, but that men do not tend to laugh in response to women's laughter. Jefferson also demonstrates that co-conversationalists will laugh in response to guttural, laughter-like tokens in speech (2010), regardless of whether or not these tokens were actually laughter implicative. She further elucidates the means by which conversationalists will work to negotiate such (mis)identified instances of laughter in an effort to reach a point of co-constructed mutual understanding of the preceding utterances as having been serious or non-serious, as the case may be. These studies laid the fundamental groundwork for our understanding of the sequential features and uses of laughter in interaction.

Building on Jefferson's legacy, researchers have continued to advance our understanding

of laughter in naturally occurring conversation (e.g., Glenn & Holt, 2013). For example, both Schenkein (1972) and Schwitalla (2001) describe a variety of interactional uses for laughter, and corroborate Jefferson's conclusions regarding careful transcript of laugh particles as a vital first step in documenting and analyzing laughter in interaction, while Hepburn and Varney (2013), Potter and Hepburn (2010), and Trouvain and Truong (2014) document the prosodic and acoustic features of laughter and interspersed laugh particles. Expanding on Jefferson's (1979) work in inviting laughter, Glenn (1989, 1991) describes how participants in conversation seek to initiate laughter subsequent to a laughable, which he defines as "whatever speakers display as laughter's object or referent" (1991, p. 140). Applying these findings to more specialized conversational contexts, Norrick (2010) analyzes the interactional function of laughter before the punchline in tellings of canned jokes, while Haakana (2002) documents the uses of laughter in medical interactions, Gerhardt (2009) analyzes laughter produced while viewing football matches, and Markaki, Merlino, Mondada, and Oloff (2010) describe the use of laughter in a professional meeting as a means of establishing group (dis)affiliations by way of an ethnic joke. An in-depth review of CA research on the sequential features and interactional uses of laughter can be found in Holt (2013a).

Turning to research examining the sequential features of utterances treated as non-serious, as opposed to the laughter following such utterances, the bedrock analyses were originally done by Sacks (1972, 1973 1974, 1978). The essential components of a sequence treated as non-serious, as defined originally by Sacks, are as follows: a) the utterance in question is followed by laughter, and said laughter is not treated as a source of trouble (e.g., through initiation of a repair sequence); b) there is no other subsequent action which could be construed as a serious response to a serious interpretation of the original utterance, and said lack of serious

next action is not treated as a source of trouble; and c) there is a jointly managed transition back to seriousness. These sequential features are defined and described in greater detail in Chapter 3.

Later studies have expanded on this foundation established by Sacks (1972, 1973, 1974, 1978). For example, Norrick (1993) provides a detailed overview of the structural and organizational features of joking in conversation, including differentiation between the sequential features of conversational joking and humor (i.e., utterances treated as non-serious) and those of canned jokes (i.e., scripted, memorized jokes with a routinized narrative structure and clear punchline). He builds on these analyses of canned jokes, in particular, in later research (e.g., Norrick, 2001; Norrick & Bubel, 2009). Similarly, Ford and Fox (2010) describe the various sequential and semiotic resources that speakers can manipulate in order to "construct laughables". On the other hand, Haddington (2011) describes a repair mechanism used in Finnish to retroactively mark an utterance as non-serious, which allows the speaker to disavow any previously understood serious interpretation if needed.

Although the prototypical response to an utterance treated as non-serious is laughter, some researchers have documented the sequential features of responses to humor other than laughter prior to the return to seriousness, for example, continued joking and echoic responses (Hay, 2001), "po-faced" responses to teasing, which consist of laughter accompanied by a serious response to the tease (Drew, 1987), and responses to failed attempts at humor (Priego-Valverde, 2009). Subsequent to the response to the utterance treated as non-serious, as noted above, there is a jointly managed returned to seriousness. One technique used in this sequential environment was first described by Schegloff (2001), who examined how the initiation and cessation of non-serious talk is negotiated by conversationalists. He demonstrates that in English, a turn-initial "no" is frequently deployed in the following turn by the participant who

made the non-serious utterance. This turn-initial "no", according to Schegloff, serves as a marker both of the termination of a sequence of non-seriousness and of that which preceded as having been, in fact, non-serious, lest one's co-conversationalists had been unaware of the import of the utterance.

Moving beyond analyses of the sequential features of laughter and non-seriousness, researchers have described a variety of interactional uses of humor in everyday conversation. In addition to his work on the sequential features of non-seriousness, Norrick (1993) offers some initial insight on this topic, with particular emphasis on how humor can be used to build rapport, to tease, and for self-presentation, as well as the use of humor to occasion the (re)telling of a story which is already known to recipients (Norrick, 2004). Similar themes are evident throughout CA research on interactional uses of non-seriousness. For example, Kotthoff (2009) notes the use of irony among children as a means of building rapport among themselves via fantasy play, as well as with the observing adult researchers by assuming an ironic "adult voice". Other researchers have noted that non-seriousness plays a key role in relationship building via "jocular mockery" amongst young men (Haugh, 2010; Haugh & Bousfield, 2012).

In terms of self-presentation, a series of studies have documented the use of humor in the workplace as a means of establishing social categories. This can take the form of boundary marking (Holmes & Marra, 2002), delineation and/or subterfuge of hierarchical power structures (Holmes, 2000), and performance of gender and/or ethnic divides (Schnurr & Holmes, 2009; Vine, Kell, Marra, & Holmes, 2009). Outside the workplace, Ervin-Tripp and Lampert (2009) describe the use of humor in (potentially embarrassing) self-disclosures, which bears strong resemblance to Jefferson's analysis of the use of humor in troubles talk (1984). Of particular note is a study by Barnes, Palmary, and Durrheim (2001), in which the means of marking an

utterance as potentially non-serious are harnessed in service of making racist statements, thereby enabling speakers to disavow, via initiation of laughter, the (potentially serious) racist import of the utterance (à la Haddington, 2011).

As illustrated by this exemplary review of the CA research of humor to date, we have learned a great deal about the sequential features of laughter and non-seriousness, and the interactional uses of the same. However, it is also clear that, in the CA tradition, humor has generally been defined in a post-hoc fashion, that is, humor or non-serious talk is defined as such by the researcher when participants in the conversation treat the previous statement as humorous or non-serious, usually by laughing. This stance is understandable, given CA's consistent avoidance of attempts to access such psychological aspects of conversation as speaker intent as a key aspect of its methodology. The goal of this methodological stance is to ground all conclusions in observations gleaned as much as possible only from the data at hand through a process of "unmotivated looking" in an effort to avoid the pitfall of incorporating the analyst's assumptions and biases into the analysis (Schegloff, 1997). Although this post-hoc definition of humor provided valuable insights into how humor is sequentially distributed and organized in conversation, this work leaves open the question of how co-conversationalists perceive that a given utterance is serious versus non-serious in order to be able to respond appropriately.

In other words, by relying on a post-hoc definition of non-seriousness, especially in instances where contextual clues are lacking, CA fails to account for how conversationalists routinely mark utterances as humorous and how co-conversationalists are able to accurately identify said utterances such that they can respond appropriately. As described earlier in the chapter, the rhetorical and psychological theories of humor presume that the content of an utterance, given sufficient logical incongruity with the preceding context, will suffice for the

detection of humor (e.g., Attardo, Attardo, Baltes, & Petray, 1994; Dynel, 2009; Norrick, 2003; Veatch, 1998). Although this is likely true for routinized, recognizable narrative forms of humor, such as canned jokes (Attardo, 1994; Norrick, 2001, 2010) and puns (Sacks, 1973; Sherzer, 1985), this leaves open the question of how participants are able to identify non-serious utterances which do not adhere to these narrative structures, such as those described by Sacks (1972) and Schegloff (2001). Indeed, anecdotally, it is not uncommon during CA-style data sessions to hear some variation of, "I can't tell from the transcript, I need to hear it again," when expert researchers are interpreting what action(s) a given utterance might be accomplishing in a given sequential context, indicating that in many cases, the content and context simply are not sufficient information to determine whether an utterance was serious or not. Before proposing a solution to this conundrum, however, let us first turn our attention to some broader issues in CA research, namely, preference and recipient design.

1.2.4. Preference and Humor

Two core findings from early research in CA describe preference for contiguity in talk-in-interaction and recipient design. Preference here refers not to individual speaker preferences, but rather to a systematic means by which certain next actions in a given sequence are structurally preferred over other potential next actions (Pomerantz, 1984; Sacks et al., 1974). In more familiar linguistic terms, preference can be viewed as a system of markedness, where preferred and dispreferred next actions can be loosely glossed as unmarked and marked next actions, respectively. Markedness is, broadly speaking, a means of characterizing linguistic forms as belonging to one of two categories: unmarked and marked. Unmarked forms are default, dominant linguistic forms which are generally viewed as more neutral variants, while marked forms are less neutral, more unusual, and/or irregular variants (Trask, 1999). For

example, in the question, “How tall are you?” the word “tall” is neutral and unremarkable, whereas in the question, “How short are you?” the word “short” stands out as an unusual semantic choice. Although ultimately the question itself is the same, in this instance, “tall” is the neutral, unmarked form, while “short” is the marked form.

This system of markedness can apply at virtually any level of linguistic analysis, from phonetics to syntax to semantics. Applying markedness at the discourse level, a preferred next action is unmarked in the sense that it is the default structurally relevant next action anticipated by co-participants in a conversation from a particular culture, while an utterance which deviates from the anticipated next action is marked or dispreferred. This system of preference is directly observable via analysis of the sequential context regularly accompanying particular categories of utterances. As has been amply demonstrated (Pomerantz, 1984; Sacks et al., 1974), preferred next actions tend to be produced by and large with little hesitation, and are often immediately contiguous with the previous utterance.

The latter feature of preferred next actions has been described as an overarching preference for contiguity in talk (Sacks, 1987). Of particular relevance for the present study is how such contiguity can be successfully accomplished through careful recipient design of utterances (Sacks et al., 1974). Recipient design has most frequently been described at the discourse level of talk-in-interaction. A classic example of this is the phenomenon of so-called pre-sequences (Sacks, 1995a), of which the clearest example are pre-invitations. In terms of preference structure, the structurally preferred response to an invitation is an acceptance (Pomerantz, 1984). In order to avoid the need for the recipient of an invitation to issue a dispreferred declination, speakers often issue a pre-invitation, in the form of questions such as, “What are you doing?” By first determining the invitee’s availability for the proposed activity

before issuing the invitation, the inviter structures the sequence such that there is a reduced likelihood of receiving the dispreferred response, i.e., a declination of the invitation (Sacks, 1995a). Such pre-invitations do the action of setting up an invitation, in that the literal meaning of the words, "What are you doing?" would normally entail a description of ongoing current activities. Yet interlocutors, regardless of what they are doing, routinely respond with, "Nothing, why?" which indicates availability for an invitation to another activity, without responding to the overt question about ongoing activities. In other words, the interaction is designed for the recipient at the discourse level to be recognizable as a jointly managed means of maintaining contiguity by encouraging a preferred response to the forthcoming invitation.

Returning to the sequential nature of non-seriousness, as described in the previous section, Sacks (1972) argues that an utterance has been treated as non-serious when it is followed by laughter and lack of serious uptake. In terms of preference structure, this is the preferred response to a non-serious utterance, while the same response would be dispreferred following a serious utterance. Sacks further proposes that there is evidence to suggest that for any given utterance, participants in a conversation must determine whether the utterance belongs to the categories "serious" or "joke". In doing so, they are able to determine the fitted, preferred response (i.e., laughter following an utterance belonging to the category "joke", or a serious response following one belonging to the category "serious"). In line with traditional notations for marked and unmarked linguistic categories, I have adopted for this study the binary category markers [+/- humor], whereby [- humor] indicates an utterance treated as serious and [+ humor] indicates an utterance treated as non-serious. Although this could be similarly notated with the privative category [humor], I have adopted the binary category [+/- humor] in order to more closely align with Sacks' (1972) proposed categories (serious, joke). These categories otherwise

adhere to the definitions proposed by Sacks (1972), differing only in notation.

In tandem with the clear evidence for a preferred response type following non-serious utterances, i.e., laughter, the same observations about the overarching preference for contiguity apply as well. For any utterance, contiguity is best maintained via accurate assessment of the [+/- humor] status of the utterance, such that the next speaker can issue the best fitted, preferred response (Sacks, 1972, 1987). By extension, the determination of [+/- humor] status is critical, precisely because a determination of [+ humor] status is potentially risky (cf. Chapter 3), in that laughter takes the place of any serious response to the previous utterance (Sacks, 1972). As Sacks notes, "the possibility of erroneous categorization by the recipient of an assertion provides the asserter with the basis for employing attempted correctives. Phrases such as 'believe me' as in 'And believe me, it's not joke because as I say, I just don't feel my life is worth living at this point' seem directed to preventing or correcting the categorization 'joke'" (1972, p. 46). Sacks subsequently argues that this suggests that any given assertion may be treated either as serious or as non-serious, regardless of speaker intent, via deployment either of the sequentially appropriate next action, or of laughter, respectively. Building on Sacks' observations, I argue in Chapter 4 that there is evidence to suggest that [- humor] is the default, unmarked category.

Although Sacks limits his argument about "erroneous categorization" (1972, p. 46) to proposing the possibility of a [+/- humor] interpretation for any given utterance, note that he also describes how speakers might employ "attempted correctives" (1972, p. 46) to clarify whether or not they are joking. In other words, speakers may explicitly mark, via various means, the [+/- humor] status of their utterance for their interlocutors, in an attempt to preempt any misunderstandings. Although sometimes the content of an utterance in its sequential context is enough information to allow successful determination of the [+/- humor] status of the utterance,

this is not always true. In contrast to instances of humor with a highly routinized narrative structure, such as canned jokes (e.g., Norrick, 1993, 2010), the content and context of the utterance alone is often ambiguous for non-serious utterances belonging to categories such as teasing and sarcasm. Marking an utterance for fellow interlocutors as non-serious and not to be treated as serious would therefore be of obvious social consequence, especially when the content and/or context of an utterance is insufficient to ascertain the [+/- humor] status of the utterance in question, thereby preventing potential misidentification.

Yet it is unclear how participants in naturally occurring conversation are able to routinely identify [+ humor] utterances, especially when responding to non-serious utterances in an ambiguous sequential environment, such that they are able to respond appropriately with laughter and lack of serious uptake of the prior turn. Given our understanding of the nature of preference structure and recipient design, and the strong influence these have on the real-time unfolding of a sequence of talk-in-interaction, it is reasonable to hypothesize that the same principles are at work in sequences containing non-serious (i.e., humorous) utterances specifically, where the preferred next response which maximally preserves contiguity is laughter and topical erasure (Sacks, 1972). The question remains, however, at which level(s) of language such recipient design is employed and to what ends. As stated above, we know from CA that recipient design certainly occurs at the discourse level (Sacks, 1995a), although it has also been claimed that it impacts all levels of interaction, from word choice (Sacks et al., 1974) to body language and gestures (Schmitt & Knöbl, 2013). Furthermore, research at the interface between CA and (socio)phonetics suggests that recipient design potentially occurs at the phonetic level as well, effecting the acoustic features of talk-in-interaction (Golato & Fagyal, 2008; Kaimaki, 2011; Pillet-Shore, 2012; Schmitt & Knöbl, 2013).

In light of these observations regarding preference and contiguity, consider the following representative excerpt containing an utterance treated as non-serious. As my collection of utterances treated as non-serious marked with the ☺ symbol grew, two key observations emerged about these utterances made with "smile voice". First, many of the utterances were not identifiable as jokes or humor solely on the basis of content and/or context. Second, despite the former observation, co-conversationalists more often than not began laughing in response prior to completion of the utterances, i.e., in overlap. Both points are illustrated in the following excerpt. As noted above, the ☺ symbol is used to indicate utterances which differ acoustically from surrounding talk and are subsequently treated as non-serious; this is seen in the following transcript as well as in all others throughout this study. The ☺ symbol is placed at the end of all lines in which the utterance fits these two criteria. It is important to also bear in the mind that the ☺ symbol does not necessarily indicate whether or not the speaker was smiling.

Excerpt 1.1.: Vorglühen bei Maximilian (Pre-Gaming at Maximilian's) (41:04 - 41:11)

```
001  Max  nun [↑sage mir liebe ↑emma, auf einer ↑skala   ]☺
          now [↑tell me  dear  ↑emma, on    a   ↑scale   ]☺
          now [↑dear emma tell ↑me,   on    a   ↑scale   ]☺
              [                                          ]
002  Emm     [·h:   huh uh     uh    huh  uh   ·h:  huh=]
              [·h:   huh uh     uh    huh  uh   ·h:  huh=]
              [·h:   huh uh     uh    huh  uh   ·h:  huh=]

003  Max  [von  eins bis zehn, (0.6)] ☺ uh::::=ja.
          [from one  to  ten,  (0.6)] ☺ uh::::=yes.
          [from one  to  ten,  (0.6)] ☺ uh::::=yeah.
          [                        ]
004  Emm  [=uhu huhuhuh   ·h:       ]
          [=uhu huhuhuh   ·h:       ]
          [=uhu huhuhuh   ·h:       ]
```

Here, Emma has begun laughing before Maximilian has completed the "humorous" utterance. However, if an utterance is to be defined as non-serious only by virtue of the fact that Emma

laughed, as per the post-hoc definition of non-seriousness described above, the question remains how she knew that laughter at precisely this moment would be perceived as interactionally appropriate in this instance without having heard the entirety of the utterance, the content of which is not overtly humorous on its face. Note, in fact, that Emma begins laughing very early in overlap (line 2), and Maximilian does not ever complete his utterance, which terminates in what appears to be an ultimately unresolved word search (line 3). The conversation picks up when another participant enters with a question unrelated to the topic at hand, in other words, there is no serious response to Maximilian's utterance, indicating that this is clear example of non-seriousness as defined by Sacks (1972).

As noted above, laughter in overlap with an utterance treated as non-serious is incredibly common. In fact, it was difficult to locate a collection of utterances treated as non-serious in which there was *not* laughter in overlap, which was a necessary prerequisite for inclusion in the present study for purposes of acoustic analysis (Chapter 3) and use in the perceptual survey (Chapter 4). On the basis of this and numerous similar examples, given that participants can and do frequently laugh prior to completion of a non-serious (or humorous) utterance, even when the content of the utterance is ambiguously humorous at best, we can conclude that participants are potentially orienting to something in the speech signal *in addition to* the content and contextual clues of the non-serious utterance proper (and potentially other aspects of interaction, such as body language, embodiment, and facial expressions).

Particularly in the case of sequentially ambiguous utterances and/or jokes with particularly sensitive content (e.g., teasing of co-present parties or culturally sensitive topics), manipulation of the acoustic stream as a means of marking phonetically an utterance as [+ humor] would be advantageous. As noted above, there is already evidence that recipient design

occurs at the phonetic level (Golato & Fagyal, 2008; Kaimaki, 2011; Pillet-Shore, 2012; Schmitt & Knöbl, 2013). However, whether this occurs, and which acoustic features might be manipulated to mark an utterance as non-serious, have not been identified in the extant literature beyond description of the acoustic features of laugh particles interspersed in speech (e.g., Hepburn & Varney, 2013; Potter & Hepburn, 2010; Trouvain & Truong, 2014).

Some attempts have been made to close this gap in the research using CA-adjacent methodologies and acoustic analyses. For example, several researchers have used naturally occurring data, as defined in Chapter 2, for acoustic analysis of humor. However, these focused on highly structured narrative sequence types (e.g., Bird (2011), who analyzes canned jokes and riddles), used data from institutional settings (e.g., Flamson, Bryant, & Barrett (2011), who use data recorded at monthly town meetings), or did not do sequential analyses to allow comparison of acoustic features across sequential environments (e.g., Hidalgo Navarro, 2011). Other researchers have offered CA-type sequential analyses, but did not use naturally occurring data. The data were instead collected, for example, in a laboratory setting (e.g., Bertrand & Priego-Valverde, 2011; Bryant, 2010, 2011) or from radio talk shows (e.g., Bryant, 2004; Bryant & Fox Tree 2002, 2005). In contrast, Kohler (2008) does use naturally occurring data and provides clear sequential analyses, however, he focuses largely on the phonetic properties of laughter in speech and how it is organized in the speech stream, arguing that "speech-smile" exists on a continuum with laughter at the other extreme.

1.3. Production and Perception of Non-Seriousness

On the basis of our current understanding of the interactional and sequential features of non-seriousness, I hypothesize that non-seriousness is routinely acoustically marked by speakers in conversation and perceived as such by co-participants. Note that failed attempts at humor and

misidentification of the [+/- humor] status of an utterance can and do happen, and must be negotiated in conversation as well, for example, by stating that one was joking in the aftermath of an inappropriate response, as observed by Jefferson (2010) and Haddington (2011). However, the very fact that these do occur and require negotiation indicates that there is likely a mechanism in place to guard against such failures by marking utterances as non-serious, even if that mechanism is not foolproof (Sacks, 1972). It has already been amply demonstrated in the CA literature (e.g., Sacks et al., 1974) that there is a preference structure for conversation in place for the avoidance of, and if necessary, negotiation of, sources of trouble in talk. Although beyond the immediate concern of the current study, this negotiation following such instances of failed and/or misidentified humor would indicate that there is likely a preference for humor to not fail or to be misidentified, which ultimately lends support to the observation that there are potentially mechanisms in place to avoid such failure and/or misidentification as a source of trouble. I argue here that "smile voice" is one of these mechanisms, in addition to the contextual and sequential clues elucidated in the extant CA work.

In Chapter 2, I describe in more detail the methodological considerations shaping the data collection procedures, the sequential and acoustic analyses, and the perceptual survey. Specifically, in light of the gaps in our understanding of the interplay between the sequential and phonetic features of non-seriousness, I argue that a combined methodology harnessing the best practices of CA and sociophonetics offers a superior method of addressing the question of how participants in conversation are able to routinely identify and appropriately respond to non-seriousness in real time. Using this dual pronged approach, I investigate a) whether and how participants in naturally occurring interactions acoustically mark utterances as non-serious, and b) whether untrained listeners can identify utterances treated as non-serious in a naturally

occurring conversation without hearing how the original participants reacted to the utterance. In other words, if "smile voice" exists as a salient feature of conversation, speaker-listeners should be able to both produce it as a means of marking humor in naturally occurring interactions, and perceive it as a cue for humor.

Chapter 3 opens with sequential analyses of the excerpts selected for use in this study. As described above, there are many interactional uses of a humorous mode of discourse (e.g., in troubles talk (Jefferson, 1984) or when making embarrassing self-disclosures (Ervin-Tripp & Lampert, 2009). However, these do not all belong to the category of non-seriousness as originally defined by Sacks (1972, 1973, 1974, 1978). In order to narrow the focus on the study, and to allow comparison of utterances from multiple speakers, all excerpts selected for this study are examples of non-seriousness as defined by Sacks. The second half of the chapter presents the results of the acoustic analyses comparing the phonetic features of serious and non-serious utterances from each speaker, in which I find that there are consistent acoustic differences between the two utterance types.

Chapter 4 presents the results of a perceptual survey distributed online. Participants in the survey were randomly assigned to two experimental groups, and either viewed video clips of the excerpts analyzed in Chapter 3, or read transcripts of the same; all excerpts were trimmed, such that the original participants' responses to the utterances were not included. Survey participants then responded to a series of questions designed to determine a) whether they believed that the participants would laugh next, or respond in some other way, and b) whether they believed that a joke had been made, or some other sort of conversational action had been done. The survey results, I argue, show that naïve speaker-listeners can in fact differentiate between utterances originally treated as serious versus non-serious by participants in naturally

occurring conversations, without having access to the conversationalists' original responses.

Finally, in Chapter 5, I offer a unified analysis of the results presented in Chapters 3 and 4. Based on the results of this study, I argue that the combined CA and sociophonetic methodology proposed in Chapter 2 represents vital progress in understanding how we produce and perceive humor in daily interaction. I ultimately conclude that, with a larger corpus, this approach could be used to glean further insights into the interplay between the sequential and phonetic features of naturally occurring conversation, not only in furthering our understanding of non-seriousness, but of human talk-in-interaction more broadly.

Chapter 2

A Methodology for the Analysis of Non-Seriousness

Excerpt 2.1.: Jaqueline und Hektor 1 (38:31 - 38:46)

```
001  H    ·ptch:: [>h::<      ]
                  [           ]
002  J            [wir könnten] nämlich über  lingUI:sten ☺
                  [we   could ] namely  about lingUI:sts  ☺
                  [we   could ]   even talk shit about    ☺

003       lästern,  = ☺
          talk shit, = ☺
          lingUI:sts,= ☺

004  H    =huh [huhuhuhuh   ]=
               [             ]
005  J         [huhuhuhuhuh?]=

006  H    =die sind    vOll     nErvig, oder¿=[huhuhuhuhuh]
          =they are complEtely annOYing,  or¿=[huhuhuhuhuh]
          =they are     sO   annOYing, right¿=[huhuhuhuhuh]
                                              [           ]
007  J                                        [huhuhuhuhuh]

008       ich  verstEhe  aber nicht  gAnz     was  sie
           i  understAnd but   not complEtely what they
           i don’t really understAnd exActly  what they

009       mAchen. aber,
            dO.   but,
            dO.   but,
```

2.1. Introduction

Research in the CA tradition has amply demonstrated that there is a preference for contiguity in conversation (Sacks, 1987). As introduced in Chapter 1 and described at length in Chapter 3, non-serious ([+ humor]) utterances are typically a) followed by laughter, and b) not followed with a response which is in any way identifiable as responding to the potentially serious import of the original utterance. Contiguity in this case, as argued in Chapter 1, is best maintained through accurate identification of the [+/- humor] status of a given utterance, followed by a fitted response. Indeed, inappropriate laughter coupled with lack of a serious response (or the reverse, a serious response coupled with lack of laughter) as a result of misidentification of the [+/- humor] status of an utterance may be treated as a source of trouble in the interaction, serving as grounds for initiation of repair (Jefferson, 2010). Such breakdowns in mutual understanding could have potentially serious consequences for the interaction beyond initiation of repair, especially in cases where the utterance references delicate topics (such as race) or currently co-present interlocutors (as in differentiating between insults and teasing).

One mechanism for managing and encouraging contiguity is through recipient design, a means by which utterances are designed to be fitted to the ongoing shifting interactional context and to the recipient(s) themselves (context-sensitive) while also adhering to shared conversational norms and preference structures (context-free) (Sacks, Schegloff, & Jefferson, 1974). Research using a CA framework has shown that non-serious utterances comprise a distinct phenomenon which has been well described (e.g., Jefferson, 1972, 1979, 1984, 2004; Jefferson, Sacks, & Schegloff, 1987; Sacks, 1972, 1973, 1974, 1978; Schegloff, 2001). Such observations are useful for analyzing the content and structure of non-serious utterances themselves, as well as their effects on the unfolding sequential structure of a given interaction.

However, CA only allows us to describe non-seriousness on a post-hoc basis. In other words, by definition, it is only identifiable after the fact when a given utterance is treated as non-serious in the manner described above, and such treatment is in turn not treated as a trouble source by any participants in the conversation. Such research fails to account for how participants are routinely able to identify a given utterance as non-serious in the first instance, such that they may appropriately respond with laughter and a lack of any identifiably serious response. It is frequently assumed that the content of an utterance in relation to the local context alone can account for this (e.g., Attardo, Attardo, Baltes, & Petray, 1994; Dynel, 2009; Norrick, 2003; Veatch, 1998), yet as argued in Chapter 1, such context-sensitive clues in the content of the utterance are frequently ambiguous at best. The highly recognizable narrative structures of certain forms of non-seriousness, such as canned jokes (Attardo, 1994; Norrick, 2001, 2010) or puns (Sacks, 1973; Sherzer, 1985), certainly aid in their identification, as does insertion of laugh particles (e.g., Hepburn & Varney, 2013; Potter & Hepburn, 2010; Trouvain & Truong, 2014). However, this does not account for how one-off non-serious utterances which fall outside these routinized structures and are not infused with laughter, such as those described by Sacks (1972) and Schegloff (2001), are identified.

The question remains, then, how participants are able to routinely perceive non-serious utterances as such in order to respond appropriately. The underlying assumption driving this study is that, given the overwhelming rarity of repair of [+/- humor] status, there must be a context-free means by which speakers may mark a given utterance for their recipient(s) as non-serious in a bid to preserve contiguity in the interaction. In the case of highly routinized non-seriousness (such as canned jokes), the recognizable narrative structure itself is one such context-free means with which speakers may mark an utterance as non-serious. A further example of

such context-free marking of non-serious utterances may include embodiments of non-seriousness, such as smiles, raised eyebrows, and “a wink and a nod”.

The focus of this study, however, is on utterances which fall outside of these recognizable narrative structures and/or the realm of signaling non-seriousness with facial expressions and body language, yet which are still nonetheless identifiable as non-serious in terms of recipient response. The working hypothesis is that one important context-free means of marking such an utterance as non-serious may be manipulation of the acoustic stream, in other words, that speakers may produce non-serious utterances with a voice which perceptibly differs from surrounding serious talk. If this is the case, then such manipulation of the acoustic stream could serve as a context-free cue marking an utterance as non-serious which, in conjunction with the local content and context, could begin to account for how participants in conversation are able to routinely identify the [+/- humor] status of a given utterance in real time.

In order to test this broader conceptualization of how participants in a conversation may accurately identify the [+/- humor] status of a given utterance, the following three hypotheses were established:

H_{1A}: Utterances which are identifiable as non-serious on the basis of how they are treated by participants in a naturally-occurring conversation are produced with acoustic features which differ measurably from surrounding serious talk from the same speaker.

H_{1B}: Absent full conversational context, participants in a survey can identify utterances which were treated as non-serious in the original conversation at a rate greater than chance would predict.

H_{1C}: When presented with acoustic data, participants in a survey can identify utterances which were treated as non-serious in the original conversation at a greater rate of accuracy than when presented with transcripts only.

In this chapter I argue that, while useful as a starting point, CA alone cannot answer the

primary research question nor test the hypotheses. Utilizing a combined approach allows us to consider the problem from a new perspective which neither tradition is able to access on its own. Given the nature of the issues raised by these hypotheses, a sociophonetic approach was selected as the complementary methodology. A sequential analysis (CA) is first used to identify utterances treated as non-serious, creating a baseline set of tokens which could be fruitfully analyzed and compared, with accepted sociophonetic methods then informing the procedures implemented in measuring acoustic features and designing a perceptual survey. In other words, sociophonetics allows us to move beyond what CA can tell us about the sequential features of non-seriousness on the basis of a post-hoc unifying definition to a fuller understanding of how participants are able to identify such utterances in real time.

An understanding of the basic *modus operandi* of these two research traditions makes clear what sort of data are required to test the above hypotheses, and how best to control for variables which may influence the experimental outcome. After describing what sort of data were necessary according to the tenets of each research tradition, I describe the specific data collection procedures as they were implemented in this study. I argue that the data collection procedures described below satisfy the basic methodological principles of each tradition, while also allowing me to address a question which neither on their own are able to address, namely, how participants in naturally-occurring conversation are able to identify non-seriousness in order to be able to respond appropriately.

2.2. Conversation Analysis

From the outset, this study has been heavily shaped and influenced by the research tradition of CA, as originally conceived and developed by Sacks, Schegloff, and Jefferson (1974). The study has its origins in my own earlier CA work, where the primary focus was on

the transition from non-seriousness back to seriousness (Trygstad, 2014). As work on that study progressed, a recurrent problem, as described in Chapter 1, was how to transcribe the acoustic variations which were perceived to routinely accompany the non-serious utterances. In early versions of that work, I initially glossed these features as "smile voice", which was marked in transcripts with a ☺ symbol. Ultimately, it became clear that CA in absence of acoustic analysis was not adequate to describe the constellation of acoustic manipulations which likely contributed to the perceptual impression of "smile voice", nor to identify what role they might play in marking [+/- humor] status.

The first batch of data which served as the inspiration for this study and was utilized in the pilot version of the survey described below was not originally collected with the intent to analyze non-seriousness. Rather, the data, beginning with telephone conversations recorded in the fall of 2007 at a large Midwestern university and continuing in Germany during the academic year 2008-09, were collected with the intent of creating a CA corpus as an end unto itself. The data were therefore collected according to the methodological tenets of CA, with an eye towards the importance of naturally occurring data and the process of unmotivated looking.

A key founding principle of CA methodology is a focus on the use of naturally occurring conversational data in analyses. Heavily influenced by the research traditions of Garfinkel (1967) and Goffman (1959, 1971), among others, Sacks (1995a, 1995b) argued that everyday conversation should be viewed as a primal site of social organization. According to Sacks but contrary to commonly held views at the time, everyday conversation is not messy or disorderly, nor is it a devolved form of linguistic performance which deviates from an idealized perfect competence. Rather, Sacks viewed conversation as fundamentally ordered and structured, describing in his lectures an interactional machinery designed to produce "order at all points"

(1995a, p. 484).

Building on this view of everyday interaction as fundamentally ordered, Sacks, Schegloff, and Jefferson (1974) proposed a system of turn-taking which accounted for the apparent "messiness" of conversation. In this way, such messiness could in fact be viewed as the fundamentally ordered results of a shared, culturally bound, yet mutually negotiated system of communication. A particularly elegant facet of their proposal argues that such a system must be simultaneously context free yet also context sensitive: context free, in that it can be deployed in virtually any interactional situation which might arise, yet context sensitive, in that it is adapted and fitted to each particular interactional situation as it unfolds in real time. Such an approach allows us to identify underlying similarities in seemingly disparate spates of talk, in other words, to analyze the actions accomplished through turns-at-talk at the meta-level, and how these interact with the content of the utterances themselves.

A key implication of viewing everyday talk-in-interaction as fundamentally ordered, and as simultaneously context free and context sensitive, is a shift away from focusing on an idealized linguistic competence. Indeed, such competency does not exist outside of experimental laboratory studies and thought experiments. Differentiating between competence and performance (Chomsky, 1965) is certainly one useful means of understanding interaction as it relates to our internal grammar. Yet in order to improve our understanding of variation in speech, we should also consider how we actually interact on a daily basis.

While some studies have attempted to access performance in the Chomskian sense, it is difficult to do so absent the use of naturally occurring conversational data. For example, if a participant in a laboratory study is asked to tell a canned joke of their choosing and recorded doing so (e.g., Attardo & Pickering, 2011; Pickering et al., 2009), this actually tells us little about

how participants do humor (or non-seriousness) in everyday talk-in-interaction. Such a study might be viewed as accessing a spontaneous performance of a joke, since the participant is not presented with a script to read and is allowed to speak freely, which will almost certainly result in "errors" in their speech, such as mispronunciations, repetitions, and hesitations, to name a few examples. However, in everyday conversation, canned jokes are not generally told in a sound booth while being recorded at the behest of a researcher.

We must differentiate here between performance in the Chomskian sense, that is, how we actually perform in everyday interactions as contrasts with our internal competence (1965), and how Sacks in his lectures described performance as engaging in particular sorts of interactions for the benefit of an outside observer, i.e., a researcher (1995b). His description bears striking resemblance to Labov's (1984) argument that any interactions taking place in the presence of a researcher creates "a formal context where more than minimum attention is paid to speech" (p. 29). The result is that the speech is not entirely natural, or to borrow his terminology, of the vernacular variety. In other words, in order to obtain a sample of speech which is as natural as possible, one must reduce to the extent possible any deviation from the natural speaking context.

Experimental designs which involve the use of a laboratory setting, therefore, tell us much about how participants perform jokes in the Sacksian sense, that is, how they perform under an experimental condition at the direction of a researcher, but they cannot tell us about how jokes are done in everyday interactions. In other words, while analysis of linguistic data in a laboratory study has taught us much about the nature of language, it ultimately falls short of revealing how participants interact with each other outside of the laboratory setting. Indeed, the disconnect between what we *actually* do in conversation, and what we *think* or *believe* we do, is striking, as demonstrated by Golato (2003). We are left then to conclude that if we want to

understand how anything is done in naturally occurring conversation, we must make precisely this the primary object of study.

This basic premise, that if we want to understand the system underlying everyday talk-in-interaction we must work with free-flowing naturally occurring conversational data, poses a particular sort of conundrum: how to produce reproducible, verifiable results. Sacks himself was also concerned with this problem, that is, how to develop a framework for doing sociological research as an empirically grounded, natural science. As he describes in his initial lectures on the topic (1995a), he originally turned to telephone calls to a suicide hotline, "for the single virtue that [he] could replay them; that [he] could type them out somewhat, and study them extendedly" (p. 622). In other words, any conclusions reached about conversation were to be grounded in data which were recorded, not derived via introspection or remembered from personal interactions, in order to permit independent verification of the results.

This fundamental methodological principle was later operationalized as a set of data collection procedures, as well as a particular way of approaching the data once collected. As a first step, Sacks argued that conversational data was best recorded for analysis and not analyzed in real time or from memory (1995a). This is because such analyses were considered unreliable at best. Indeed, it has been demonstrated that our memory of the details of our interactions is generally incomplete and lacking in detail (e.g., Stafford, Burggraf, & Sharkey, 1987). Furthermore, it is virtually impossible to recall and attend to the full details of conversations, such as pauses, gaps, hesitations, mispronunciations, and false starts, among others, all of which have been demonstrated as having profound implications for talk-in-interaction (e.g., Pomerantz, 1984). Therefore, it is vital that all data be recorded and available to the researcher for multiple hearings, such that they may attend to even the minutest details of the interaction.

In addition to recording conversational data for future analysis, any data analyzed under the CA framework must consist of interactions which did not take place at the behest of a researcher or in a laboratory, rather they should ideally be interactions which would likely have taken place for reasons completely independent of the needs and interests of the researcher. In order to determine whether this is the case, interactions which may be considered naturally occurring and therefore suitable for analysis should satisfy the so-called "dead social scientist" test (Potter, 2002). This test is a sort of benchmark used to determine whether a given interaction may be considered naturally occurring. In order to perform the test, the researcher must pose to themselves the following question: "[W]ould the data be the same, or be there at all, if the researcher got run over on the way to work?" (Potter, 2002, p. 541). If the answer to this question is "No," then the interaction cannot be considered naturally occurring, and while such an interaction may be useful for analysis for any number of purposes, it cannot be used for analysis of naturally occurring conversation as such.

Consider the following example: A researcher invites a pair of volunteers to a laboratory to participate in a study. Even if the plan were to ask the participants to interact "normally" for a period of time by engaging in a free-flowing conversation, such an interaction would not and furthermore could not proceed if the researcher failed to arrive at the laboratory, meaning that it cannot be considered naturally occurring. On the other hand, if the exact same plan were to take place in a participant's home or other location of their choosing with a friend or relative, chances are very high that the participants would proceed with their plans to socialize regardless of the presence or absence of the researcher. While they certainly might think it strange that the researcher did not arrive as planned, there would be little reason not to proceed to interact, and such an interaction can therefore be considered naturally occurring, according to the "dead social

scientist" test.

Once recorded, in order for the results of the analysis to be available to and replicable by fellow researchers, Sacks argued that the data should then be meticulously transcribed. Although not trained as a phonetician, Sacks' co-analyst, Jefferson, sought to transcribe the original recordings in as much detail as possible (Ochs, 1979), albeit impressionistically, a point to which we will return later in the chapter. In this way, a reader could at a minimum read the transcribed conversation and partially access the nuanced features of the original interaction, enabling them to verify or even dispute the analyst's findings as they saw fit (Sacks, 1995a). Any resultant differences in opinion regarding a given analysis on the basis of the published transcript were ideally to be resolved by listening to the original data recordings wherever possible, an ideal which has been actualized on a grander scale with modern technologies which enable easier sharing of recordings (Schegloff, n.d.). Indeed, analysts have insisted that transcripts must not be considered data in and of themselves (e.g., Heritage & Atkinson, 1984; Ochs, 1979), rather one should consistently return to the original recordings throughout the analytical process.

A further advantage of transcribing the data is that the process itself forces the researcher to engage with the interaction at an extreme level of detail. This in turn encourages a process of "unmotivated looking" (Sacks, 1995b). As noted above, Sacks argued that we should avoid to the extent possible approaching the data with a particular preconceived notion of what we might find or what sorts of problems we might solve (1995a). To reiterate his point from above, this again is precisely why he turned to telephone calls to a suicide hotline in the first instance: because they were already recorded and available to him, not because they were particularly interesting or because they answered a particular question he had in mind. Rather, through unmotivated looking and noticing of interesting structures over time, the researcher should

discover what we can learn from the data independent of what problems or assumptions we might otherwise bring to bear. This approach, in Sacks' view (1995a), can lead us to far more interesting results and discoveries than if we have from the outset predefined the question, which in turn places certain limitations on the potential answers we might find.

Returning to how these methodological principles come to bear on the present study, I initially became interested in non-seriousness purely by happenstance. In the fall of 2008, I attended a short course taught by Professor Celia Kitzinger at the University of York. While working together on a spate of talk culled from my original data corpus, she drew my attention to the following snippet:

Excerpt 2.2.: Andreas und Pete (09:55 - 10:19)

```
001  P    e- e- es war ↑unfassbar    wie viel dieser mensch
          i- i- it was ↑unbelievable how much this   person
          i- i- it was ↑unbelievable how much this   person

002       gekotzt hat.=das  war=
           puked  had.=that was=
          puked      .=it   was=

003  A    =huh:[::::      ·h::]=
          =huh:[::::      ·h::]=
          =huh:[::::      ·h::]=
004  P         [              ]
               [wirklich toll ]=
               [ really  great]=
               [awesome       ]=

005  A    =es war so wie    mister creosote, ☺
          =it was so like   mister creosote, ☺
          =it was just like mister creosote, ☺

006  P    huh:↑jauhhuh  es war     ähnlich,  >es war< wah::=
          huh:↑yesuhhuh it was     similar,  >it was< wah::=
          huh:↑yahuhhuh it was kinda like that,>like< wah::=

007  A    =den eimer  [wah:]::[:h:]
          =the bucket [wah:]::[:h:]
          =the bucket [wah:]::[:h:]
```

```
                          [     ]  [   ]
008  P                    [>·h<]  [°h°]
                          [>·h<]  [°h°]
                          [>·h<]  [°h°]

009       (.4)

→  010  A   <ja  ↑das ist=[↑coo:l.>] ·h:: ↑JA?  (.4) ich habe,
            <yes ↑that is=[↑coo:l.>] ·h:: ↑YES? (.4)  i  have,
            <yes ↑that is=[↑coo:l.>] ·h:: ↑YES? (.4)  i  have,
                          [         ]
011  P                   =[°·h::°   ]
                         =[°·h::°   ]
                         =[°·h::°   ]

012  A   (.1) wann war das , vorgestern       haben wir
            (.1) when was that, before yesterday have  we
            (.1) when was that, the day before yesterday we

013       ehehuh: kam  es dann auch nochmal zu einer, ·h::
          ehehuh: came it then also  again  to   a  , ·h::
          ehehuh: it also came again to a           , ·h::

014       es wurd- >also es wurde geraucht.< ja?
          it  was- >well it  was  smoked  .< yes?
          it  was- >well it  was  smoked  .< you know?

015       (.6)

016  P    ↑mm?
          ↑mm?
          ↑mm?
```

In this excerpt, the utterance in line 5 "*Es war so wie Mr. Creosote* / It was just like Mr. Creosote," is treated by both participants as non-serious, in that both participants laugh in response and engage in further hyperbolic non-serious utterances comparing the volume of vomit produced by Pete's friend to that produced by the character Mr. Creosote during the famous skit in *Monty Python's The Meaning of Life* (Goldstone & Jones, 1983). The utterance in line 6, "*Ja, es war ähnlich* / Yeah, it was kinda like that," should not be mistaken for serious agreement with the utterance in line 5, because treating as serious the suggestion that somebody had produced a

volume of vomit similar to Mr. Creosote would inevitably involve disagreement instead. Rather, this should be viewed as alignment with the non-serious work done by Andreas in line 5.

In managing the transition from non-seriousness back to seriousness, note that Andreas deploys a turn-initial "*Ja* / Yes" twice in line 10. While working with this excerpt, Prof. Kitzinger pointed out that this token is similar to the turn-initial "no" deployed in a similar sequential environment as described by Schegloff (2001), in that it belonged to the same category of speech, namely, agreement tokens. Of interest, however, was that it had the opposite valence (a positive agreement token as opposed to a negative disagreement token). Having noted this particular turn-initial "*ja* / yes" token after a non-serious utterance as a potential point of interest, I later noticed that it occurred frequently in this sequential environment (Trygstad, 2014). This finding aligns well with Betz' (2017) description of turn initial "*ja*" as a discourse marker which does not necessarily indicate an affirmative response, but rather can serve as a marker of discontiguity with the prior turn. In the case of non-seriousness, such discontiguity could serve as an indicator of the discontinuation of non-seriousness, and in turn as a marker of the return to seriousness, as described by Schegloff (2001).

As this collection of sequences containing utterances treated as non-serious grew, two closely related observations stood out. First, without listening to the recording again, it was at times difficult to tell from the transcript alone (that is, from the content and context) how a co-conversationalist was able to identify an utterance as potentially non-serious in order to be able to respond appropriately. A corollary observation was that utterances treated as non-serious were often made with a voice which differed perceptibly from surrounding serious talk made by the same speaker, lending credence to my analytical judgment of the utterances as non-serious. Eventually, I began to suspect that the acoustic features of a given utterance might interact with

the content and context of the utterance as a means of marking it for the recipient as non-serious and aiding in its identification.

As an initial means of investigating the hypothesis that the acoustic features of an utterance could be manipulated by a speaker to mark an utterance for their recipient as non-serious, I set out to transcribe in more detail how utterances treated as non-serious differed from surrounding serious talk. Struggling with how to transcribe these utterances using standard CA transcription conventions, I settled on marking the utterances with a ☺ symbol (as in line 5, above), a stop-gap solution which revealed little about the acoustic features of the utterances beyond the analyst's own subjective impressions of a non-serious or "smile" voice. This problem with accuracy in transcription led directly to this project, making it truly a product of unmotivated looking.

Ultimately, as argued above, a CA approach by itself was not enough to address the question of how participants identify non-seriousness as such in the first instance, nor how to transcribe the constellation of acoustic variations which often accompanies such utterances. Since its inception, we have learned much from CA as a direct result of its unique approach to data collection and analysis, yet the methodology also poses several limitations which the field must confront in order to continue improving our understanding of the nature of naturally occurring conversation. With regard to this study in particular, significant limitations of CA were the focus on sequence-level analyses and the quality of the audio data.

Originally, CA focused primarily on analysis of sequence organization with only very superficial consideration of the acoustic features of the utterances in question. Despite the insistence noted above that analysts listen back to the original data rather than relying on transcriptions (e.g., Heritage & Atkinson, 1984; Ochs, 1979), very little was done to

systematically analyze acoustic features of naturally occurring speech in traditional CA studies. Phonetic descriptions were often done impressionistically, rarely measured acoustically. Indeed, many of the classic findings were based on data which was of such poor quality that acoustic analyses would have been nearly impossible. Given that we now have more sophisticated tools at our disposal, e.g., superior recording equipment and phonetic analysis software such as PRAAT, it is imperative that we continue to do more to consider the acoustic features of talk-in-interaction in a systematic, empirically verifiable way.

Indeed, important inroads have been made in recent years in measuring the acoustic features of the speech signal and considering how the results of phonetic analyses may come to bear on the sequential analysis. Many researchers (e.g., Couper-Kuhlen & Selting, 1996; Golato & Fagyal, 2008; Kaimaki, 2011; Wennerstrom, 2001) have issued calls for increased attention to these issues, arguing that we must consider the role of prosody in talk-in-interaction, and how it comes to bear on its sequential features, while others have offered detailed descriptions of how we can better implement and utilize the results of such mixed methodologies (e.g., Ogden, 2015; Walker, 2017).

In similar fashion to CA findings writ large, while the post-hoc analysis of non-seriousness revealed much about how non-seriousness may be identified and described sequentially (e.g., Jefferson, 1972, 1979, 1984, 2004; Jefferson, Sacks, & Schegloff, 1987; Sacks, 1972, 1973, 1974, 1978; Schegloff, 2001), we know little about the acoustic features of non-seriousness (with the key exception of prosodic analyses of laugh particles in non-serious utterances; e.g., Hepburn & Varney, 2013; Potter & Hepburn, 2010; Trouvain & Truong, 2014), nor what role (if any) acoustic signals may play in how speakers are able to identify non-seriousness in the first instance and respond in a sequentially appropriate manner. In order to

address this question, we must move beyond what CA alone can tell us about the shape of sequences containing non-seriousness and consider other variables in the interaction, such as the acoustic features of utterances treated as non-serious. Given the hypotheses presented above and following the models proposed by Couper-Kuhlen and Selting (1996), Golato and Fagyal (2008), Kaimaki, (2011), and Kohler (2008), among others, a more in-depth acoustic analysis is an important first step.

Yet an omnipresent problem with the use of data collected "in the wild" is the effects of external noise pollution on the quality of the audio data. Although most acoustic features are clear enough for a sequential analysis, poor acoustic conditions can lead to difficulties with accurately measuring the acoustic signal with phonetic software. However, in order to advance our understanding of the role that acoustic and perceptual phonetics play in conversation, we must find ways to make better quality audio recordings despite environmental limitations while also obtaining naturally occurring data outside of a laboratory setting.

The data collection procedures described below, I argue, satisfy the methodological principles of CA: the data were recorded using audio-visual equipment, they can be classified as naturally occurring, and no assumptions were made prior to collection about what the solution to the transcription problem posed by "smile voice" might be. Additionally, while certainly it was hoped that utterances treated as non-serious would occur, no special prompts were given to the participants beyond being encouraged to interact with each other as they normally would. At the same time, several additional controls were put in place in order to address the methodological limitations described above. Specifically, conversations were limited to dyads, recording locations were selected and practices were implemented to improve the quality of the recorded acoustic signal, and laboratory data were collected from all participants for future comparative

analysis.

Once recorded, the data were first transcribed and subjected to a traditional sequential analysis. The results of this analysis, presented in Chapter 3, indicate that the utterances culled for use in the acoustic analyses and perceptual experiment were treated as non-serious by both participants according to the CA definition. In this way, the groundwork of the study lies squarely in a sequential analysis informed by CA of non-serious utterances, yet we turn to methodological tools from the tradition of sociophonetics in order to push that analysis further than would be possible with CA alone.

2.3. Sociophonetics

To better understand the role of acoustic and perceptual phonetics of non-seriousness in naturally occurring conversation, we turn to the tools offered by sociophonetics as a means of expanding on the results of the sequential analyses presented in Chapter 3. Sociophonetics has traditionally placed much emphasis on acoustic analyses. In recent years, however, a renewed focus on perception has expanded the sociophonetic repertoire (e.g., Hay, Warren, & Drager, 2006; Thomas, 2002). In line with current trends in the field (Jannedy & Hay, 2006), this study considers both production and perception of non-serious utterances through a series of acoustic analyses and a classic perceptual experiment. Distinguishing it from other sociophonetic studies, however, I analyze utterances culled from naturally occurring conversational data.

Sociophonetics has its origins in the merger of sociolinguistics, which focused on linguistic variation and social stratification, with experimental acoustic phonetics, which documented the acoustic properties of the inventory of human speech sounds. This focus on discovering and documenting the properties of human speech made necessary an experimental setup in a quiet laboratory setting in order to isolate testable variables to the extent possible (such

as might be introduced by noise pollution or by variation in recording equipment and procedures). The seminal works of William Labov marked a major shift in the field. In a series of studies, Labov (e.g., 1963, 1972) documented how the acoustic variation of speech correlated with shifts in formality and usage. In other words, acoustic measurements could be shown to vary not just from language to language and dialect to dialect, but also according to social context.

Labov's studies were groundbreaking, not only for his unique use of statistical methods which allowed documentation of speech variation and change in progress, but also for his use of data outside of a laboratory setting. However, they also initially relied on impressionistic assessments of what was said (e.g., his 1972 survey in New York City department stores), where Labov would write down what he heard after the interaction took place. Although such techniques could not take into consideration the potential effects of perceptual limitations on the part of the researcher, the findings were nonetheless instrumental in shedding light on the potential value of naturally occurring data by viewing variation in speech as a resource on which speakers draw rather than as a confounding variable.

Later studies attempted to overcome the limitations of Labov's work by applying the acoustic measurement techniques which were still being developed in the realm of acoustic phonetics. This enabled a more fine-grained phonetic analysis of the data, however, there was still a tendency to rely on interview data and carrier phrases (e.g., Labov, 1984). Such studies certainly allow for the isolation of variables and comparison of speakers across a variety of social categories, however, they ultimately can tell us little about the phonetics of naturally occurring conversation, as argued above. Indeed, as Labov himself notes in his telephone survey of New Yorkers (1966), it was most fortuitous when speakers were interrupted during a phone

interview because such side interactions produced more natural speech with the greatest dialectal influence.

Concurrently, a growing number of researchers have begun to investigate the role of perceptual phonetics. As it has become clear that even trained transcriptionists are very limited in their ability to overcome their own perceptual biases, the necessity of accurately measuring the acoustic signal as opposed to making an impressionistic assessment has been placed in stark relief (Kerswill & Wright, 1990; Thomas, 2002). Incidentally, such findings dovetail nicely with the CA perspective noted above, i.e., transcripts should be a considered a tool for analysis rather than data as such (Heritage & Atkinson, 1984; Ochs, 1979).

Much has been learned from the application of tools from acoustic and perceptual phonetics to questions about language variation and change, however, few studies to date have attempted to use the same tools on naturally occurring data as defined by CA. This study seeks to fill this gap by subjecting excerpts culled from unaltered naturally occurring conversational data to a fine-grained acoustic analysis and a classic perceptual experiment. Specifically, by using the tools offered by sociophonetics to measure the acoustic features of utterances treated as non-serious and to test whether variation in the speech signal of the same utterances is salient to untrained (i.e., non-linguist) listeners, we may begin to address the question of how participants in everyday talk-in-interaction overwhelmingly are able to identify non-seriousness. While it could be argued that using naturally occurring data introduces unnecessary variables which must be accounted for in the analysis, when we as analysts choose to utilize data recorded in a laboratory, the laboratory itself may in fact be one unaccounted variable for which we must control by increasing the degree to which the presence of a third party observer could influence the interaction (i.e., the classic observer's paradox). In other words, if we wish to understand the

nature of everyday language use, we must make this the object of study by reducing the extent possible the perceived presence of an observing researcher in an unfamiliar condition.

2.4. Data Collection Procedures

In the following, I describe the procedures used to collect the data presented in Chapters 3 and 4 of this study. The collection procedures described here were selected and designed to test the hypotheses derived from the research questions presented in Chapter 1 in accordance with accepted practices in the fields of CA and sociophonetics, as argued here. All interventions with human subjects were approved by the Institutional Review Board of the University of Wisconsin-Madison and were conducted in compliance with recognized ethical standards for research on human subjects. Sample informed consent forms can found in Appendix 2.1.

2.4.1. Production

The production branch was designed to test one of the hypotheses guiding this study. The central hypothesis of the study, H_{1A}, restated from above, was as follows: Utterances which are identifiable as non-serious on the basis of how they are treated by participants in a naturally-occurring conversation are produced with acoustic features which differ measurably from surrounding serious talk from the same speaker. If the corollary null hypothesis (that such utterances do not differ acoustically from surrounding serious talk) could be rejected, it would serve as evidence that participants may mark utterances as designedly non-serious in the speech stream using a constellation of acoustic features or "smile voice". While beyond the scope of the present study, this raised the question whether this hypothesis would be best tested with naturally occurring conversational data or with laboratory data; I hypothesized that there would be differing results for each data type.

To this end, three types of data were collected: audio-visual recordings of naturally

occurring conversations suitable for analysis using CA, a hybrid form of data consisting of audio-visual recordings of naturally occurring conversations recorded in a laboratory setting, and audio recordings of sentences read aloud in a laboratory setting suitable for acoustic analyses. The collection of these three types of data allowed for conducting both CA and acoustic analyses of instances of seriousness and non-seriousness and comparing both inter- and intra-speaker results, which addresses the first hypothesis (H_{1A}), while also laying the groundwork for future comparative work on non-seriousness which tests the methodological argument that naturally occurring data will produce results which differ from laboratory data.

2.4.1.1. Audio-Visual Data Collection

The conversational data for this study were gathered during the academic year 2012-2013 in Germany. Upon arrival in Germany and after obtaining Institutional Review Board (IRB) approval from the University of Wisconsin-Madison, contact was established with potential participants. Participants consisted mostly of friends and acquaintances of the researcher, as well as contacts located through a campaign of e-mails, Facebook postings, and flyers (Appendix 2.2.) distributed over the campus of the University of Potsdam (Universität Potsdam) and in Berlin, Germany. All participants were asked to complete a series of three recordings, although not all participant pairs completed the entire series.

All volunteers who contacted the researcher were asked to pick a friend with whom they felt comfortable interacting freely and openly for an extended conversation. This friend could be a close friend, a partner, an acquaintance the volunteer had not seen for a long time, a co-worker, or any other person of their choosing. It was only indicated to them that the person should be somebody whose company they enjoyed and with whom they felt they could carry on a conversation at length comfortably, and who was willing to consent to participate in the study.

Once a second volunteer had been located in this manner, an appointment was agreed upon for the first recording session.

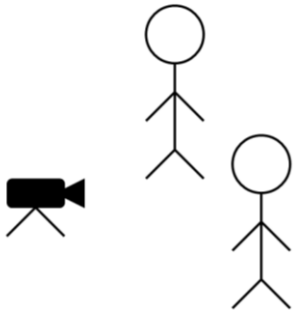
The first recording was completed at a location of the participants' choosing, and was designed to be in line with the methodological tenets of CA described above. The participant pair was asked to choose a time and location that was convenient for them, as long as the location was quiet (to allow for good quality recordings for acoustic measurements) and likely to be free of interruption (for example, by roommates, children, or other individuals not consenting to participation in the study). It was indicated to the participants that they may engage in an activity of their choosing during this time, such as eating, working on a project, or playing a game, as long as said activity was quiet (noisy games and listening to music, for example, were discouraged) and involved active engagement in conversation (passive activities such as reading together or watching a movie were also discouraged). Such interactions were all therefore likely to proceed, regardless of my presence or absence. In this manner, the experimental setup satisfied the requirements of the so-called "dead social scientist" test (Potter, 2002) described above.

The number of participants in each conversation was purposely limited to two. This was done to control for the influence of the number of participants on the interaction, as well as to reduce the potential for overlapping talk (which was of particular importance for the survey portion of the study described below). Furthermore, all participants were asked to agree to complete all recordings together as a pair once a partner had been selected. These controls were put into place because while they increased the comparability of the data from a sociophonetic perspective, said data could still be considered naturally occurring talk-in-interaction from a CA perspective.

All participant pairs elected to complete the first recording session at the home of one of the participants (or of both, in the case of one pair consisting of a couple). Upon arrival at the recording site, the researcher assisted the participants in selecting a suitably quiet location in the home and arranging appropriate seating which was conducive to audio and video recording. A typical setup took the form of a triangle, with the points consisting of the two participants and the camera equidistant between and in front of them. An example of this setup is illustrated in the diagram below, Figure 2.1.

Figure 2.1.

Typical Camera Setup



In addition to the camera, a laptop and two external microphones were used to the record the interaction. The camera was placed on a tripod with an external microphone attached. This microphone was placed approximately equidistant between the participants. The laptop was placed on a nearby table or chair, and an external microphone was attached and placed between the participants, generally at closer range than the microphone attached to the camera. This allowed for a superior audio recording to be made on the laptop with Audacity for use in acoustic analyses, while also taking audio through the camera which was synced with the video and could be used for CA analyses of the data. In order to achieve a wide enough camera angle to allow the participants to sit at a comfortable distance from each other, it was necessary to place the

camera, and its microphone as well, at some distance from them. The microphone attached to the laptop, on the other hand, could be placed much closer, capturing superior quality audio data.

Once the camera had been set up and the participants had been guided through the informed consent process, I turned on the recording devices and exited the room or home, retiring to another room or nearby location previously agreed upon in consultation with the participants. It was indicated to the participants that they should attempt to converse for a minimum of one hour, but that they could continue to interact for as long as they pleased, even if the memory card in the camera ran out, with the understanding that the audio recording on the laptop would continue for an indefinite period of time for all intents and purposes. They were also told to contact the researcher via telephone when they wished for the recording devices to be turned off at any time and for any reason, even if they had not yet conversed for an hour, in which case I would then return promptly and switch off the devices.

After switching off the recording devices and ascertaining whether the participants were comfortable with continuing their participation in the study, an appointment was made to complete the second and third recordings in a joint session. Some participant pairs were unable or unwilling to complete the second recording session, usually as a result of the prohibitive costs of travel for participants not already living in Berlin or its suburbs. In these cases, the third recording session (as described below) was generally completed at the conclusion of the first recording, and no further data were collected.

For the joint second and third recording sessions, each pair of participants was invited to travel to the Institute for German Studies (Institut für Germanistik) of the University of Potsdam in Potsdam, Germany. The recordings all took place in an office in the Institute. The office served as a substitute to a laboratory or sound booth, which was unavailable. Participants were

provided with light snacks and drinks during the second recording session in order to ease the discomfort of interacting in an unfamiliar and relatively restricted environment. They were positioned at a small, square table in the triangle configuration shown in the diagram below. The participants were seated in a configuration as similar as possible to the naturally occurring experimental condition, such that the two participants again formed an equilateral triangle with the camera. The table was placed at an angle to the camera in order to enable the participants to see each other while seated at a comfortable distance and to allow the camera to capture as much of their nonverbal communication as possible.

As in the naturally occurring experimental condition, the participants were recorded using both a video camera with external microphone and Audacity on a laptop with external microphone, resulting in a video recording with synced audio of slightly lower quality and a high quality audio recording of the interaction. The equipment used was identical in both experimental conditions. However, due to the location, there was increased control of noise pollution. The resultant data were therefore generally cleaner and of better sound quality than in the naturally occurring condition, given that the acoustic features of each location varied considerably, as did the level of ambient noise pollution.

In addition to increased control over ambient noise the use of a laboratory-type setting for all participants allowed for the control of a further variable in the study. Specifically, the recording environment was different for each participant pair during the first recording phase, which was a necessary artifact of the choice to utilize the method of data collection required to perform CA analyses, wherein the researcher records events that would normally take place in whichever location participants have decided to gather, with the result that each location is unique. While this method of data collection allows us to examine language in its native

environment and better approximate its natural state, it also introduces a host of complex environmental variables which are difficult to a) identify and b) tease apart from the results.

In order to control for these (largely unknown) environmental factors and their (potential) effect on the data, the second recording was completed in the laboratory, as described above. While the data still consisted of naturally occurring, free-flowing conversation (participants were allowed and even encouraged to engage in conversation according to their own interests), the experimental conditions did not satisfy the requirements of normal CA data collection procedures, in the sense that the participants would have been unlikely and unable to proceed with their plans if the researcher had not arrived at the previously arranged time and location.

During the second recording session, I remained outside the door of the office. This allowed the participants to easily attract attention at any point by simply opening the door, should they become uncomfortable with continuing the recording session, as well as allowing easy monitoring of the door to prevent intrusion during the recording sessions. On one occasion, such intrusion occurred despite all precautions, resulting in the elimination of a brief segment of data from consideration in the study due to privacy concerns for the individuals who had not consented to participation. At a time of their choosing, the participants opened the door and signaled that they wished to conclude the recording session, at which point the recording devices were promptly switched off. At the conclusion of the second recording session, it was again ascertained whether the participants wished to continue participation in the study.

Although not all participant pairs completed the second recording session, all participants completed the third recording session. For participants who could not or would not travel to Potsdam, this third recording session was completed immediately following the first recording session in a quiet and secluded area of their respective homes. For all other participants, the

third recording session took place in the office cum laboratory immediately following the second recording session. Other than location, all participants completed the third recording session in the same manner as described below.

For the third recording session, participants were presented with a series of sentences to read aloud in a normal speaking voice to represent the serious condition (see Appendix 2.3.). Immediately afterwards, they were asked to read the same series of sentences aloud "as if they are funny or with a smile" ("*als ob die lustig sind bzw. mit einem Lächeln*") for the non-serious condition. The sentences were carefully chosen to be ambiguous without their surrounding context. In other words, they could have been entirely serious utterances or entirely non-serious, a determination which was impossible to make without context. Additionally, seven of the ten sentences were culled from data previously collected in Darmstadt, Germany, during the academic year 2008-09. That is, these were actual sentences which had been uttered in the course of naturally occurring conversations. These sentences were selected in order to replicate naturally occurring German language and conversation as closely as possible. Three of the sentences had originally been uttered as non-serious with smile voice, one had been uttered with smile voice but with serious content and intent, and the rest had been uttered as serious with no smile voice. The remaining three sentences were invented examples which were verified to be grammatically correct by native speakers of German prior to implementation of the experimental condition.

The recordings were made by first asking one participant to leave the room so that the sentences would be new on sight to each participant and also to reduce the discomfort associated with reading aloud in front of an audience. Using the external microphone and Audacity, the participants were recorded on the same laptop used during the recording of their conversations

during the first and second recording sessions. Participants were asked to read each sentence aloud three times in a row, that is, the first sentence was read three times, followed by the second sentence, and so on. Once all ten sentences had been read three times in the serious condition, participants read the same sentences again in the non-serious condition, following the pattern of reading each sentence three times before proceeding to the next. This allowed the participants to become more comfortable with the sentences and also allowed for multiple collections of the same data point in the instance that a sentence were garbled or otherwise not clearly recorded on the first attempt. Participants were unaware that they would be reading the sentences a second time in the non-serious condition. They only became aware of this after the sentences had been fully recorded in the serious condition and they flipped the card to reveal the second set of instructions with the sentences reproduced below. Once the sentences had been read aloud three times each in both the serious and non-serious conditions, the participant was asked to leave the room while the second participant made their recording.

These sentences were intended to replicate an experimental laboratory condition, where participants are asked to read words and phrases aloud to test various features of speech. The third recording session therefore represents ever-increasing control of the variables in comparison to the first and second recording sessions. Not only was the environment controlled for, but the object of study as well, i.e., the sentences provided by the researcher. This allowed for the establishment of a baseline for comparison, both inter- and intra-speaker, of acoustic measurements made using the naturally occurring data.

This increasing control of variables in the second and third recording sessions allows us to compare the results of the phonetic analyses of the first recordings to those of the second and third recordings in order to determine what effect, if any, environment and context could have on

the phonetic features of non-seriousness. In addition, the elimination of these variables also allows for stronger inter- and intra-speaker comparisons of such phonetic features. Such comparison will allow us to test the arguments put forth by proponents of CA methodology that conversation is best analyzed in its natural environment and by proponents of acoustic phonetics that such data are too messy, that the variables cannot be controlled, and that experimental data can tell us as much (or more, due to increased control of confounding variables) about phenomena such as smile voice. Given that laboratory data are significantly easier to collect, measure, and analyze, it is important to determine in which instances laboratory data are sufficient to address the research question, and in which instances naturally occurring data are truly necessary. Although these questions are beyond the scope of the present study, such considerations point to an important avenue for future analysis of the data of the sort initially collected for this study.

The excerpts culled from these data for acoustic analyses were selected on the basis of preliminary CA analyses which will be explicated in Chapter 3. In short, excerpts were selected which contained instances of non-seriousness according to the CA definition (the utterance was followed by laughter, there was no serious response, and such response was not subject to repair) and which were uttered in the clear (i.e., not in overlap with laughter or other talk). The latter criterion was necessary in order to allow for acoustic analyses as well as for use in the perceptual survey. Finally, all excerpts of non-seriousness and seriousness also fit the selection criteria for inclusion in the perceptual survey, as described below. This allowed for comparison of the results of the production and the perception branches of the study.

2.4.2. Perception

The perception branch of the study was designed to test the hypothesis that the

constellation of acoustic features which accompany non-seriousness are salient and meaningful to participants in conversation and not an artifact of the researcher's biases. Because it is not possible to access the immediate, ongoing interpretations of participants in live conversations, a perceptual experiment was selected to access the interpretations of uninformed, non-expert (i.e., non-linguist) observers. The survey was designed to collect data which would allow us to test Hypotheses H_{1B} (absent full conversational context, participants in a survey can identify utterances which were treated as non-serious in the original conversation at a rate greater than chance would predict), and H_{1C} (when presented with acoustic data, participants in a survey can identify utterances which were treated as non-serious in the original conversation at a greater rate of accuracy than when presented with transcripts only). Rejection of the corresponding null hypotheses (that participants could NOT identify which utterances were treated as non-serious, and/or that access to the acoustic signal did NOT correlate with increased accuracy in doing so) would serve as evidence that variation in the speech signal may play a role in how participants are able to identify the [+/- humor] status of an utterance.

2.4.2.1. Survey Data Collection

The survey for the perception study was designed and tested in two distinct phases. In the first phase, an initial design, as described below, was created and piloted in the summer of 2012. The results of this initial pilot study were presented at a graduate student colloquium at the University of Potsdam in Potsdam, Germany in the fall of 2012. The changes that were made to the pilot design of the study were informed by the results of the study itself, as well as by feedback during the colloquium and discussion of the design with Dr. Jürgen Trouvain at the 17th Working Conference on Research on Conversation and Discourse (17. Arbeitstagung zur Gesprächsforschung) in Mannheim, Germany in the spring of 2013.

The pilot version of the survey was designed and run using a commercially available online survey tool, Survey Gizmo (http://www.surveygizmo.com/). The decision to use an online survey platform instead of an in-person tool, such as the survey capabilities available through PRAAT, was made due to the lack of native speakers of German available to complete the survey in person in the United States. Although certainly there are a number of disadvantages associated with an online survey, such as technical problems or failure to follow instructions (a problem especially pertinent to this survey design, as will be discussed below), the advantage of remote access to an exponentially larger population far outweighed these potential problems.

Survey Gizmo was selected over other commercially available internet-based products (such as Survey Monkey) and the survey host offered by the University of Wisconsin-Madison (Qualtrics) because it included all features required for the purposes of this experiment. The free educational version of Survey Gizmo allowed for an unlimited number of survey items with accompanying audio-video material and item randomization (on the second version of the survey). All three of these features together were not available with the other services; unlimited survey items, accompanying audio-visual material, and item randomization were only available with the paid version of Survey Monkey, which was beyond the means of the researcher, and accompanying audio-video material was not supported by Qualtrics.

Initially, the intent was to collect data on survey participants' immediate reactions to audio-video material containing instances both of seriousness and non-seriousness. Specifically, they would be asked to view video of a naturally occurring conversation amongst native speakers of German and answer two questions about the video: what just happened in the video, and what will happen next. As this portion of the survey was designed to access participants' immediate

"gut" reactions to the videos in order to access to the extent possible their real time impressions, they were asked to view the videos only once before answering the questions. Although it was technically impossible to prevent participants from viewing the videos multiple times, the decision was made to proceed with an internet-based survey design in order to reach a larger audience, on the assumption that most participants were likely to follow the instructions. The full text of this spontaneous reaction survey can be seen in Appendix 2.4. (Survey A1).

In addition to these data, I quickly determined that it would be useful to know the participants' reasons for why they selected the response that they did from the perspective of the participants themselves. To this end, three additional open-ended questions were added: two asking why they selected the response they did on the questions about what just happened and what will happen next, and a third pertaining to what aspects of the interaction stood out to them in responding to the survey items. The full text of this considered response version of the survey can be seen in Appendix 2.4. (Survey A2). However, in order to access participants' thoughts on why they selected their responses and which features stood out to them, it would be necessary for them to view the data multiple times and reflect on their responses. This would eliminate the possibility of collecting their immediate initial reactions to the data, rendering it impossible to collect perceptual data which mimicked live responses as closely as possible, which was a primary goal of the perceptual branch of the study.

Therefore, the final survey design contained two sections, essentially combining the two designs into one larger survey. Participants were first asked to view the videos only once and answer the two "gut" reaction questions on a spontaneous reaction survey (Appendix 2.4., Survey A1). After viewing the videos once each and answering the questions, they proceeded to the second part of the survey, the considered response portion (Appendix 2.4., Survey A2).

Here, they were presented with the same videos as viewed in the first part and with accompanying transcripts of the audio. They were again asked what happened in the video and what would happen next, in addition to the three open-ended questions asking why they responded as they did. For this part of the survey, they were instructed to view the videos as many times as deemed necessary and to read the accompanying transcript. This provided them with more information about the interaction and more opportunity to reflect on their responses.

The excerpts for the pilot version of the perception branch were selected from two distinct data sets originally recorded for and used in previous CA work (Trygstad, 2014). The first data set consisted of phone calls recorded in the fall of 2007, while the second set, consisting of audio-video data, was recorded in 2008-09 in Darmstadt, Germany. Consisting of naturally occurring everyday conversations among native speakers of German, the data were recorded according to the methodological tenets of CA, and were therefore suitable for CA analysis. Such data are necessarily and inherently messy in comparison to laboratory data, potentially introducing a host of variables (some of which I attempted to control for in the final version of the study, as described below). However, as discussed earlier in the chapter, the decision to use such data was made because perceptual data collected in response to laboratory data would tell us little about how participants perceive non-seriousness in naturally occurring conversation, effectively introducing the variable of using a test condition which does not address the research question. In order to test the hypothesis that participants could differentiate between seriousness and non-seriousness in everyday conversation, it was necessary to use conversational data in the perception measurement tool, i.e., the survey.

The excerpts included in the pilot survey were selected on the basis of analyses completed for the earlier study (Trygstad, 2014). In order to test what effect, if any, the use of

audio only versus audio-video data might have on the results, two instances of non-seriousness were selected from each data set (two from audio-only telephone calls and two from audio-visual in person conversations), for a total of four. The excerpts were selected according to the following criteria: a) each was an instance of non-seriousness according to the CA definition (i.e., the utterance in question was followed by laughter and treated as non-serious by co-participants in the conversation); b) the laughter did NOT occur in overlap with the non-serious utterance (allowing for trimming as described below); and c) the utterance was made using a voice that was perceived by the analyst to differ acoustically from surrounding talk from the same speaker. Serious excerpts were selected at random from the data pool. The only selection criteria were the following: a) the utterance was NOT an example of non-seriousness according to the CA definition; and b) the utterance was made in the clear (i.e., not in overlap with any other talk to allow for trimming as described below).

Once selected, the non-serious excerpts were trimmed using iMovie software exactly at the end of the utterance, but prior to any laughter. Serious excerpts were trimmed at a transition relevance place (or TRP) (Sacks et al., 1974), such that laughter in principle could have been a potential response at that point. In this way, survey participants were prevented from knowing what actually occurred after the utterance of interest (i.e., they had no knowledge of the audience reaction). Their judgments on what occurred in the excerpt and what would happen next were instead based solely and entirely on the utterance itself, in addition to the minimal context provided prior to the utterance in order to aid in understanding the content.

An issue with the pilot design of the perception survey concerned randomization of the excerpts. At the time of the initial design, it was not possible to truly randomize the excerpts. In order for the survey program to display the questions in a unique, randomized order to each

participant, it would have randomized the excerpts across the entire survey rather than within the Survey 1 and Survey 2 question sets. This would have had the highly undesirable outcome of mixing the spontaneous "gut" reaction type questions with the considered response questions, eliminating the possibility of collecting these types of data separately and in the correct sequence. The decision was reached to number the excerpts, and use a random number generator to assign a new sequence to the numbered excerpts. This new sequence was maintained across both surveys (i.e., the first excerpt was the same on both surveys, the second was the same, etc.). In this way, the excerpts were displayed in a randomly determined order, but this order was fixed for all participants, and therefore was not truly randomized. It was necessary to proceed at the time with this pseudo-randomization for expedience in collecting data for the pilot study, on the assumption that the randomization problem could likely be resolved at a later date for the full study, which indeed was the case.

The pilot study was distributed to native speakers of German living in the United States. Although care was taken to avoid recruiting linguists and/or conversation analysts in the full study (especially those familiar with the study), many non-CA linguist contacts participated in the pilot study. This allowed them to advise the researcher on the design of the study, in addition to collecting pilot data. Given that the ideal participant would have been a non-linguist native speaker of German currently residing in a German-speaking country, this represented a convenience sample which would not overlap with the limited number of contacts available to the researcher who were eligible to participate in the full study. This was an important consideration, as participation in the pilot version would necessarily eliminate one's eligibility to participate in the full study. Participants were unaware of the object of study (i.e., perception of non-seriousness) in order to avoid influencing their responses. However, in order to fulfill the

requirements of the Institutional Review Board to reveal withheld information upon completion of participation in the study, they were made aware of this at the end of the survey during debriefing (the full text of which can be viewed in Appendix 2.1.), rendering it impossible to participate in both the pilot and final versions.

On the recommendation of Dr. Jürgen Trouvain (personal communication, March, 2013), the final survey design introduced a second experimental condition to serve as a further control. Specifically, Dr. Trouvain suggested that although I was testing the question concerning whether participants could identify non-seriousness at a rate greater than chance with the then-current survey design, I was neglecting to address the question of whether or not the speech signal played a role in how participants perceived non-seriousness. To address this concern, it was determined that a second survey would be created and the participants would be divided into two experimental conditions, creating a computer-controlled blind study.

The second version of survey was designed to closely mimic the first version, with two sections of the survey (spontaneous reaction and considered response) and identical data excerpts, with one crucial difference: the first portion of the survey for one half of the participants contained only transcripts with no accompanying audio-visual material. Participants in Experimental Group B would be asked to answer the survey questions based only on the transcript material, and then asked to answer the same questions again on the second part with the aid of the accompanying audio-video material, again, as with Group A, providing them with more information about the interaction on Survey 2 and more opportunity to reflect on their responses. In summary, Section 1 of the survey for Experimental Group A contained audio-video material only and for Group B, transcript material only, while Section 2 of the survey was identical for both experimental groups (A and B). The full text of the Group B version of the

survey can be viewed in Appendix 2.4. (Surveys B1 and B2).

The inclusion of the second experimental group allowed me to examine what effect, if any, access to the audio-visual material versus access to the content of the conversation only (in transcript form) had on participants' accuracy in identifying instances of non-seriousness. Certainly in order to fully isolate the variable of the speech signal, it would have been necessary to harvest the audio from the video, and create additional experimental conditions to account for the effect of including video in the survey. However, in order to fully control for this variable, this would have resulted in the creation of nine experimental groups in theory. In practice, only five of these nine groups would have been appropriate permutations of the experiment. As described above, in order to allow for the considered response condition in the second half of the survey, participants were provided with more information than in the first half. In other words, Experimental Group A received audio-visual data, followed by audio-visual data with transcript; Group B received transcript data followed by audio-visual data with transcript. This represents a consistent movement from less information to more information, which eliminates at the outset four of the nine theoretically possible permutations, as indicated by the red squares in Table 2.2.

Table 2.2.

Potential Survey Condition Permutations			
Survey 1	**Survey 2**		
	Transcript Only (TO)	**Audio + Transcript (AT)**	**Audio-Visual + Transcript (AVT)**
Transcript Only (TO)	n/a	TO → AT	TO → AVT (B)
Audio Only (AO)	n/a	AO → AT	AO → AVT
Audio-Visual Only (AVO)	n/a	n/a	AVO → AVT (A)

In order to fully test the effect of audio only versus audio-visual data on the survey results, it would have been necessary to include five total experimental conditions, as indicated by the yellow and green squares in Table 2.2., above. However, the addition of four experimental conditions to the original setup (AVO → AVT) would have increased the number of participants necessary to achieve a large enough sample size beyond the feasibility of the means and funds available to the researcher. The decision was therefore made to limit the survey to the inclusion of only one additional experimental condition, for a total of two experimental conditions (AVO → AVT and TO → AVT, as indicated in the green squares in Table 2.2. above). This allowed us to test what effect, if any, the inclusion of audio-visual data had on participants' accuracy in differentiating instances of non-seriousness from seriousness without overtaxing the logistic capabilities of the researcher.

The decision was made to include video in the survey on the basis of two important factors. First and foremost, given that the goal of the survey was to determine how participants perceive non-seriousness in naturally occurring conversations, I did not want to deny them access to data to which the original conversationalists had access, i.e., visual cues. Decontextualizing the audio from the video therefore would have *introduced* an additional variable, namely what effect denying access to visual information present to the original conversationalists would have on participants' responses to the survey items, bringing us back to the need to control for these variables and to include all five of the experimental conditions described above.

Perhaps the most important factor in making this decision, however, was that in the pilot version of the study there was no effect on ability to differentiate between utterances treated as serious versus non-serious based on whether participants viewed video data or listened to audio-

only telephone conversations. Given these results, it was concluded that inclusion of video was unlikely to greatly affect the outcome of the experiment. This certainly does not control perfectly for the visual cues variable; in conducting future research in this area, it would be fruitful to distribute a survey with five experimental conditions to a much larger pool of participants in order to fully explore these variables and their effects on perception of non-seriousness.

With the final survey design in place, excerpts were selected for inclusion in the survey from the newly collected audio-visual data pool described above, i.e., from the data collected for the production branch of the study in Germany during the academic year 2012-2013. As a result of feedback received during evaluation of the pilot phase, a set of controls were put into place during data collection. Specifically, during the pilot phase, the number of participants in the conversational excerpts viewed in the survey ranged from two to six. Especially for non-linguists unaccustomed to viewing and interpreting naturally occurring conversation, this introduced a potentially problematic variable in that there were a number of voices speaking, and it was at times difficult to focus on a given speaker and isolate instances of speech, which stood in stark contrast to the less chaotic dyadic telephone conversations. For this reason, in addition to those described above, I collected data only in dyads. The result was data which were cleaner and less confusing for participants to view and interpret. In addition, the quality of the audio of these original excerpts was poor, and therefore better equipment was obtained in order to allow for better acoustic appreciation of the speech signal by survey participants. In this way, an attempt was made to control for the inherent "messiness" of naturally occurring conversational data (as compared to laboratory data), such that we could eliminate some of the variables introduced by this data type in a survey environment.

In addition to placing increased controls on the data pool itself, I also attempted to control for two variables introduced by the speakers of the utterances of interest. Specifically, care was taken to select non-serious utterances and serious utterances from a set of only four speakers. In other words, once the four non-serious excerpts had been selected, serious excerpts from the same speaker were located in order to eliminate the variable of speaker variation, creating a set of corresponding pairs of four unique speakers. This stands in contrast to the pilot version, where the non-serious utterances were all made by speakers different from those who made the serious utterances, for a total of eight unique speakers. Additionally, the serious excerpts for each speaker set were selected from the spate of talk immediately prior to that which contained the non-serious utterances. In this way, the serious excerpts were in some sense prosodically and topically contiguous with the non-serious excerpts, controlling for the influence of content and topic shift on participants' judgments of seriousness versus non-seriousness.

The selection criteria for the non-serious excerpts were similar to those utilized during the pilot phase. The original three criteria were retained, as stated above: a) each was an instance of non-seriousness according to the CA definition (i.e., the utterance in question was followed by laughter and treated as non-serious by co-participants in the conversation); b) the laughter did NOT occur in overlap with the non-serious utterance (allowing for trimming as described below); and c) the utterance was made using a voice that was perceived by the analyst to differ acoustically from surrounding talk from the same speaker. However, an additional fourth criterion was introduced on the basis of initial analyses and literature review. Specifically, as described in Chapter 3, given that the focus of the study was on the acoustic features of instances of non-seriousness as opposed to other uses of "smile voice", the decision was made to limit the survey excerpts to utterances which were not followed by any identifiably serious uptake by the

co-participant in the conversation, representing increased precision in defining non-seriousness over the pilot version of the study. The selection criteria for serious utterances did not change from the pilot study, with the exception of the controls described above in selecting utterances immediately prior to the non-serious utterances. Finally, all excerpts utilized in the perception study were also analyzed in the production branch of the study. Therefore, all excerpts selected met the criteria for use in the acoustic analyses.

As in the pilot phase, once selected, all non-serious excerpts were trimmed using iMovie software precisely at the end of the non-serious utterance, such that the response (i.e., the laughter) could not be heard by survey participants. Similarly, all serious excerpts were trimmed at a TRP (Sacks et al., 1974), again, so the response could not be heard by survey participants. In this way, it was not possible for participants to determine what sort of turn had occurred on the basis of audience reaction, encouraging them to attend to the utterance itself in responding to the survey items.

The final step in the design phase was to integrate randomization at three main points in the survey. The first point of randomization was a branching feature. This randomly assigned 50% of survey participants to Experimental Group A and the remainder to Experimental Group B. The purpose of randomization at this point was to create a computer-controlled blind study, such that participants were aware neither that they had been assigned to a group, nor to which group they had been assigned, while the researcher is aware that they were assigned to groups and which group was assigned. A double-blind study design was unnecessary as it was deemed unlikely that researcher bias would affect the outcome of the study, given that participants did not have any direct interactions with the researcher and that the researcher did not become aware of which experimental group had been assigned until after the survey had been completed and

submitted.

The second point was randomization of the excerpts. As noted earlier, it was not possible to randomize the audio-visual data excerpts during the pilot phase because the randomization would have randomized all pages of the survey. The result was that it would have been impossible to ensure, for example, that the informed consent page displayed first or that all the questions from Survey 1 (spontaneous reaction / transcript only conditions) displayed prior to any of the questions from Survey 2 (considered response condition). These technical difficulties were overcome in the final design, such that the two sets of excerpts and corresponding survey items were organized into two groups, which were then randomized within the groups, but not across the entire survey. The result was that the informed consent, demographic information questions, equipment testing, instruction, and debriefing pages all had a fixed position in the survey, while the Survey 1 and Survey 2 excerpts and survey items were randomized within their respective groups. This randomization of the excerpts was implemented to mitigate any response bias associated with displaying survey items in a fixed order to all participants.

Finally, a third point of randomization was included by randomizing the survey response options. Two of the survey items for each excerpt asked participants to choose among several suggested options, or to select "Other" ("*Etwas anderes*") and fill in an explanation blank. These response options were randomized amongst themselves, with the "Other" option remaining static at the bottom of the list. The purpose of this point of randomization was two-fold. First, it was intended that participants consider all the options available to them for each survey item. By forcing participants to read the options carefully each time, rather than assuming they knew what the list contained, it was hoped that they might be more likely to select the most fitting option. In addition, this point of randomization was also intended to reduce survey fatigue by providing

a slightly novel set of stimuli for each survey item.

With these final major design changes in place, the survey was distributed to a small group of native and non-native speakers of German living in the United States for a final test phase. Participants in this test phase were asked to schedule an appointment with the researcher to meet and complete the survey in person. The purpose of this exercise was to ensure that the survey had been programed properly and to identify and address any issues with clarity and/or functionality in the survey. A number of small problems with the HTML coding and Flash video embedding were discovered during this process, in addition to identifying several typos and points where instructions where unclear. For example, it was noted during this phase that nowhere in the survey was it clear that participants should NOT attempt to click the back arrow in their browser winder, as this would terminate the survey; a note was then added to the bottom of each page indicating that participants should only click on the "Forward" ("*Weiter*") button in the survey and never use the browser buttons. Each of the minor issues identified during this final test phase were corrected, with no major changes required to the final design or text of the survey.

After the final test phase was complete and all changes had been implemented, the survey received wide distribution beginning in January 2014, with secondary waves in April 2014 and in January 2015. This was accomplished via an electronic campaign targeting native speakers of German living in German-speaking countries. Participants currently residing outside of Germany, Switzerland, or Austria were included in the study, but were not the focus of active recruitment. There was no attempt to target non-linguists specifically, although linguists already familiar with the specific goals of the project were asked to distribute the survey link but not participate themselves. Given that demographic data were collected (Appendix 2.5.), it was

possible to ascertain what affect, if any, formal linguistic education had on survey responses. The electronic campaign consisted of a series of emails, postings and private messages on Facebook (http://www.facebook.com), and postings on LinguistList (http://linguistlist.org/) which were directed at native speakers of German and non-native speakers of German known to have native speaker contacts. Recipients were asked to a) participate in the survey (if they were native speakers), and/or b) forward the message to any native speakers they knew. Included in the emails and private Facebook messages were two attachments containing flyers for physical distribution (Appendix 2.2.). One of the flyers was a tear-away design, intended to be posted in a public location, while the other was a quarter sheet leaflet to be distributed by hand. Both versions of the flyer contained a scannable QR code which directed mobile users to the mobile version of the survey. Individuals currently living in German-speaking countries were asked to distribute these flyers if possible.

Collection of these perceptual data allows us to compare the results of the production branch of the study with the interpretations of non-linguist observers by considering the acoustic measurements and the non-linguist perceptions of identical spates of talk. In this way, it was possible to determine whether the constellation of acoustic features associated with non-seriousness is salient and meaningful to participants in a survey at a rate greater than chance would predict. By extension, we may then consider whether these are also salient and meaningful to participants in everyday conversations.

2.5. Conclusion

CA has made a strong case for using data which are collected directly from the context we wish to study. With that in mind, for present purposes, if our goal is to understand the parameters of non-seriousness in naturally occurring everyday conversation, we must base our

analysis on such data. Interview data or participants recording canned jokes in a laboratory setting will not suffice. This view is further supported by findings from Labov (1966), which documented great variation in how participants spoke when answering interview questions as opposed to during brief interruptions during the interview.

However, as a methodology, CA alone is limited by the sort of questions that we can address. In particular, due to the use of impressionistic methods of acoustic analysis (Ochs, 1979), any conclusions which are based on how an utterance is made phonetically are subject to the perceptual biases of the analyst (Kerswill & Wright, 1990; Thomas, 2002). Yet on the other hand, the reliance on interview data and carrier phrases in sociophonetic studies leaves us unable to generalize any results gleaned from a more fine-grained acoustic analysis to everyday interactions.

This study is unique in that it heeds the call to incorporate a more objective acoustic analysis with our sequential analysis (e.g., Couper-Kuhlen & Selting, 1996; Golato & Fagyal, 2008; Kaimaki, 2011; Wennerstrom, 2001). It harnesses the strengths of each tradition in an attempt to push the boundaries of what we can learn from each. In the manner described above, the utterances selected for inclusion in the study were culled from naturally occurring conversations. The conversations were recorded in dyads and in a quiet environment with high quality equipment in order to produce recordings with audio quality suitable for acoustic analysis and for use in the perceptual survey. The results of a sequential analysis (as described in Chapter 3) reveal that all utterances were treated as non-serious during the interaction, and such treatment was not subject to repair. Although the utterances differ in terms of specific content, they are sequentially similar and can therefore be considered comparable from an experimental standpoint for use in both the acoustic and the perceptual experiments. By drawing on the

methodological tools offered by both CA and sociophonetics, we can begin to address the question of how participants in conversation are able to identify non-seriousness as such in real time.

Chapter 3

Production of Non-Seriousness

Excerpt 3.1.: Anderson Cooper's CNN Town Hall with James Comey (10:06 - 10:27)

001	**AC**	·ptch >uh i wanna ask you< something else that
002		was in the memos °it was revealed° the ↑prEsident
003		(.) uh in a conversation with you, >according to
004		you,< made a jo:ke about how journalists might
005		↑change ↑their ↑minds ·h: about revealing their
006		↑sOUrces if they spent a night in jail, ·h:: you
007		lAUghed in respOnse, what's funny about jailing
008		journalists.
009		(0.4)
010	**JC**	nothing=i ↑lau:ghed to make it ↑clear that was a
011		joke.
012		(1.2)
013	**JC**	>it was my way< of ↑cOmmUnica:ting, °·h° he ↑MUST
014		be kidding.

Note. Transcribed interview of Comey (2018). Video retrieved April 26, 2018, from CNN website, https://www.cnn.com/videos/politics/2018/04/26/james-comey-town-hall-full.cnn. Transcription is my own, including any errors or deviations from original audio.

3.1. Introduction

Subjective analysis of utterances treated as non-serious by participants in natural conversation suggests that such utterances often are produced in ways distinct from surrounding serious utterances. Specifically, speakers often utilize a "voice" which is perceptually distinguishable from surrounding talk, but which cannot be easily captured as, for example, simple changes in pitch or intensity. CA has revealed much about the sequential structure of non-serious utterances, while on the other hand, studies utilizing sociophonetic approaches have rarely systematically analyzed non-seriousness in naturally occurring conversational data. Such studies have instead generally harvested data from a variety of linguistic contexts, for example, jokes recorded in a laboratory (e.g., Anolli, Ciceri, & Infantino, 2000; Attardo & Pickering, 2011) or culled from scripted television dialogues (e.g., Attardo, Eisterhold, Hay, & Poggi, 2003; Purandare & Litman, 2006; Urios-Aparisi & Wagner, 2011).

As argued in Chapter 2, neither of these approaches used in isolation can explain the unique acoustic features of naturally occurring non-serious utterances, nor how these might come to bear on the ability of one's interlocutor to identify accurately the [+/- humor] status of a given utterance in order to be able to respond differentially and appropriately. In consideration of these initial acoustic impressions and methodological concerns, the first hypothesis presented in Chapter 2 was defined as follows:

> **H_{1A}:** Utterances which are identifiable as non-serious on the basis of how they are treated by participants in a naturally-occurring conversation are produced with acoustic features which differ measurably from surrounding serious talk from the same speaker.

The corresponding null hypothesis is that non-serious utterances do not differ measurably from surrounding serious utterances from the same speaker. If this can be rejected, an important

underlying question can then be addressed in the analysis. Specifically, it could be the case that:

a) speakers manipulate some or all acoustic features in the same way (e.g., all speakers increase their pitch);
b) speakers manipulate some or all acoustic features in entirely different ways (e.g., speakers change their pitch, with some increasing and others deceasing, the only unifying factor being some measurable change from their individual baseline pitch); or
c) speakers manipulate some or all acoustic features in a variety of ways which brings the speech signal to converge on a normative non-serious "voice" (e.g., speakers change their pitch in such a way that causes the non-serious utterance to have a pitch characteristic of non-serious utterances across many speakers, whether by increasing, decreasing, or not changing their pitch at all, depending on their individual baseline pitch).

All these scenarios could then potentially vary based on such variables as gender and regional varieties of German, an additional possibility which is further explored in this chapter.

In order to test this hypothesis and advance our understanding of non-serious utterances in everyday conversations beyond their sequential features, this study consists of two main branches, namely, a production and a perception branch. The production study described in this chapter opens with CA-style sequential analyses, followed by acoustic analysis. Each excerpt culled from the data for use in the manner described below was first analyzed using CA in order to verify its status as a non-serious utterance, and to locate a serious counterpart from the same speaker in contiguous speech immediately previous to the non-serious utterance. The excerpts were then subjected to acoustic analysis in order to identify potential differences between speakers' serious and non-serious utterances. These same excerpts were later utilized in a perceptual experiment, as described in Chapter 4, allowing comparison of the acoustic analyses

with participants' responses on a survey. Consideration of the results of the sequential analyses in conjunction with those of the acoustic and perception experiments allows us to approach a fuller understanding of the interplay between sequence and prosody.

All excerpts selected for further analysis contained utterances which were treated as non-serious by the interlocutors, a determination made on the basis of the sequential analyses presented in the first half of this chapter. Once the data had been collected in the manner described in Chapter 2, the recordings from the participants' first recording sessions (the naturally occurring conversations) were first annotated at all points of audible laughter. This step served as a primary filtration to identify potential instances of non-seriousness. From this large collection, all instances where a co-participant in the conversation laughed in overlap with the speaker were immediately eliminated from use in the current study for two reasons: it would have been impossible to a) measure accurately the acoustic features of the utterances, and b) include such instances in the perceptual survey without revealing the participants' responses. This step significantly reduced the size of the collection of potential instances of non-seriousness which were usable for this study, as participants in a conversation frequently begin laughing at the end of non-serious utterances in slight overlap, as discussed in Chapter 1. Although certainly these utterances were examples of non-seriousness in need of further analysis, they were not usable for this study.

This smaller collection of utterances was then subjected to a preliminary sequential analysis. As argued in Chapter 1, the goal of this step was to eliminate uses of "smile voice" which did not truly belong to the category of non-seriousness, such as when it accompanies serious utterances in order to soften criticisms or express embarrassment. Similarly, utterances which were intrinsically humorous to the participants (as evidenced by laughter), but which

could not be sequentially shown to belong to the category of non-seriousness because they were treated as serious in some manner by the interlocutors, were eliminated from consideration. Due to their highly routinized narrative structure, so-called "canned jokes" (Attardo, 1994) were also excluded. Finally, although many are identifiable as instances of non-seriousness from a sequential perspective, puns, word play, witticisms, and other related phenomena unfortunately are generally much too short in duration to conduct acoustic analyses of the nature described in the latter half of the chapter, and therefore were also excluded from use in the present study. While these potential uses of "smile voice" could (and should) be considered in future work in this area, especially as they compare to the variety of non-seriousness analyzed here, the goal of the present study was to define the acoustic features of one type of utterance only, i.e., non-serious utterances, as a means of gaining initial analytic purchase on an analysis of the acoustic features of a single category of use of "smile voice".

The remaining collection consisted of a series of utterances which were treated as non-serious by the participants in the conversation, as revealed through sequential analysis. As stated above, of these utterances, only four could be selected for inclusion in the perceptual branch of the study, and the same four would be analyzed acoustically in the production branch. In selecting the final four utterances, several factors were taken into consideration. Utterances which required a large amount of context to be included in the excerpt in order to comprehend them were eliminated from consideration, as were utterances which were mumbled or otherwise garbled. Such utterances would have been difficult or impossible to analyze acoustically, and also would have been unduly taxing for participants in the survey branch, contributing unnecessarily to survey fatigue. Similarly, any utterances where ambient noise external to the conversation severely disrupted the audio signal were also eliminated. While all of the excerpts

in this penultimate collection again were solid examples of non-seriousness from a CA standpoint, and are valuable data points in the study of the sequential features of non-seriousness, ultimately, these external factors rendered them unusable in the acoustic and/or perceptual experiments described in the present study.

The elimination of such utterances yielded a final collection of six potential excerpts. These were then presented as a collection at a student colloquium at the University of Potsdam (Universität Potsdam) in Potsdam, Germany in the summer of 2013. On the basis of feedback received during the colloquium and in subsequent discussions, the four examples of non-seriousness which were least ambiguous, i.e., those which most clearly exemplified the sequential features of non-seriousness, were selected for inclusion in the final study.

On the basis of the sequential analyses presented below, I argue that these excerpts contain utterances which were clearly treated as non-serious by the participants in the conversation. The same utterances, with corresponding serious utterances from the same speakers, were then subjected to a series of acoustic measurements. The results of the acoustic experiments suggest that speakers can and do manipulate the acoustic stream from their normal baseline patterns when making non-serious utterances. However, the specific constellation of manipulated acoustic features comprising an individual's "smile voice" may differ from speaker to speaker. Although a greater number of tokens are needed to demonstrate statistical significance for the selected parameters, the results suggest trends which demonstrate the value of this methodological approach and warrant the pursuit of a larger-scale study with a similar experimental design.

3.2. Conversation Analysis and Non-Seriousness

By way of introduction to the detailed analysis of each excerpt, I first provide an

overview of the features of the sequential context of non-seriousness. Each of the excerpts presented below can be considered examples of utterances which were treated as non-serious by both participants in the original conversation. While each excerpt naturally varies greatly in terms of its overt features, they all follow a similar underlying structural pattern which can be considered characteristic of an utterance treated as non-serious, specifically: both participants laugh subsequent to the utterance, such laughter is not treated as a source of trouble in need of repair, there is no response to the utterance which treats it as serious in import or action, such lack of serious response is not treated as a source of trouble in need of repair, and there is a clear return to seriousness jointly managed by both participants in the conversation.

All conversations used for this study are dyadic, as detailed in Chapter 2, therefore, the structural paradigm below is only presented in terms of dyadic interactions. Especially as concerns deployment of laugh tokens and the potential for initiation of repair, conversations with greater than two speakers could potentially have a very different sequential structure (Sacks, 1974), as may be the case, for example, when two of three participants treat an utterance as non-serious, while a third does not laugh. However, in order to gain initial analytic purchase on the acoustic features of non-seriousness, for the methodological reasons stated in Chapter 2, only dyads were used for the present study. Therefore, only instances where *both* participants treated an utterance as non-serious were considered a non-serious utterance for the purposes of this study, although one can certainly envision a scenario where some participants in an exchange treat an utterance as non-serious while others do not, yet such an utterance could still be characterized as having been treated as non-serious. As a result, limiting this study to dyads also simplifies in a certain sense the issue of whether an utterance can be considered non-serious on the basis of how the participants in the conversation treated it. That being the case, similar

patterns are certainly found in conversations of three or more participants, although with some slight revisions, in particular to how laughter can be meaningfully deployed (Sacks, 1974).

Furthermore, I am not making claims about speaker intent. From a CA perspective, we cannot make a claim about whether or not an utterance is intended as [+/- humor], only about how it was treated by the participants in the interaction. Treatment of an utterance as non-serious tends to have the sequential features described below. Intent, in contrast, cannot be shown sequentially (e.g., Heritage, 1990). Analyzing prosodic cues may give us some insight into how participants might encourage certain treatment over others (i.e., as non-seriousness), but ultimately, there is no way to access psychological intent via any means currently available to us as researchers.

With above caveats in place, the most overt interactional cue that an utterance is being treated by participants as non-serious is laughter subsequent to or in partial overlap with the utterance. In reference to canned jokes, a highly narrative form of non-seriousness, Sacks notes that the unproblematic, default response sequence to the completion of a joke's punchline "consists wholly of laughings" (1974, p. 374). This can be also be extended and considered the unproblematic, default response that can be deployed by speakers to treat an utterance as non-serious in general, not just in the case of canned jokes. Norrick (1993) even goes so far as to label laughter as a second pair part in the adjacency pair "joking and laughter" (p. 23).

That is not to claim that any and all laughter treats an utterance as non-serious, however, as has been well-documented, for example, in Jefferson's (1984) work on laughter initiated in troubles talk by the troubles-teller. Rather, laughter can be seen in a given interaction as *potentially* treating an utterance as non-serious, although the entire sequential context must certainly be taken into consideration, both in terms of what preceded the laughter (i.e., to what

the laughter can be seen as being responsive; Sacks, 1974) and what comes next (i.e., how the laughter itself is treated subsequently and whether a serious response is forthcoming, as described below). Laughter can be treated, for example, as a source of trouble and subject to initiation of repair (Jefferson, 2010). If repaired, we can observe that, while the recipient member of a dyad may have initially treated an utterance as non-serious, if the speaker initiates repair on the [+/- humor] status of the utterance by not laughing, denying that laughter on their part was intended (Jefferson, 2010), or seeking a serious response (as below), then we cannot claim that *both* parties in the interaction have treated the utterance as non-serious, as the status of the utterance has now been observably repaired to serious by the speaker.

Laughter may be initiated by either the recipient or the speaker, or may occur in virtual overlap. If laughter occurs in overlap with the non-serious utterance itself, then by default it has been initiated by the recipient; if it occurs as infusion of laugh tokens in the utterance itself, then it has been initiated by the speaker. When it occurs immediately after the non-serious utterance, with little or no gap, it may be the case that both parties laugh virtually simultaneously, with no clear initiator. On the other hand, if laughter does not occur immediately, with a resultant gap of silence, then the speaker may invite recipient laughter by laughing themselves (Jefferson, 1979). In this way, the speaker makes clear that the utterance is to be treated as non-serious, and, if the invitation to laughter is accepted by the recipient, then the utterance has, up to this point, potentially been treated as non-serious by both parties.

If the prior sequential context suggests an utterance may be non-serious, both parties in the conversational dyad laugh (regardless of who initiated said laughter), and such laughter is not subject to initiation of repair, then we have a first indication that an utterance has been treated as non-serious. The next important sequential feature to which the analyst must attend is whether

or not a serious response is made to the utterance in question subsequent to the laughter. As Sacks argues in his work on suicide hotlines, "If an assertion is heard as 'joke,' then a single device [i.e., laughter] is appropriate without regard to what the assertion might be seen to propose. […] The use of that response operates to (a) close the sequence which the assertion began and (b) to detopicalize the assertion" (Sacks, 1972, pp. 44-45). In other words, by laughing, the recipient treats an utterance as non-serious, and in doing so, renders sequentially irrelevant any potentially serious import or action that an utterance might otherwise have been seen to have had or taken. The sequence is effectively closed by the laughter (Jefferson, 1972), and in the case of non-seriousness, takes the place of a response which could identifiably be seen to respond to a potentially serious interpretation of its import or action.

This key sequential feature, that laughter closes a sequence (Jefferson, 1972), is best illustrated by Sacks (1972). In describing calls made to a suicide hotline, many callers described loved ones indicating a wish to die. Such callers at times stated that they had not taken the assertion of suicidality seriously, and some had even laughed in response, only to then discover, upon the death of their loved one, that the assertions of suicidality were, in fact, gravely serious. Yet Sacks argued that, if one were to believe that an assertion of suicidality were serious, then the sequentially appropriate next actions would not involve, as he describes, "[…] conversation, but hospitalization, as its appropriate response. If the assertion is heard as 'serious,' then the matter thereby raised may not be disposed of until its proposer has died" (1972, p. 45). In short, a serious response to the [- humor] assertion of suicidality entails significant emotional upset, as well as disruption of all other activities in order to summon assistance. Given that a serious assertion of suicidality is, generally speaking, unimaginable for most, and that to assign a [- humor] status to such an unimaginable assertion would require a significant, taxing response, it is

unsurprising that many might wish to deny that such an assertion is even remotely serious. However, in treating an assertion of suicidality as non-serious by, for example, laughing, the sequence is closed. To the immense relief of the suicidal individual's interlocutors, no further serious action is needed, or, as Sacks writes, "To detopicalize is then to erase" (Sacks, 1972, p. 45).

Failure on the part of the recipient to respond to the serious action of an utterance can, of course, be a source of trouble in interaction, and therefore subject to repair. If, subsequent to laughter and lack of serious response, the original speaker initiates repair by seeking a serious response, then we again have an instance where we cannot analyze the utterance as having been treated as non-serious by both parties in the dyad. Instead, by seeking a serious response, the [+/- humor] status of the utterance has again been repaired to serious by the speaker. Alternatively, even if the speaker laughs, a recipient may decline to laugh and instead offer a serious response, as in the case of troubles talk noted above (Jefferson, 1984). In other words, while the speaker may display troubles-resistance by laughing, the recipient may offer a serious response, ultimately treating the original utterance as serious.

However, it can be the case, for example, that an interlocutor finds an utterance *inherently humorous* and displays this with laughter, yet does not treat it as non-serious. In this scenario, while both participants in the dyad may laugh subsequent to the utterance, if the recipient then offers serious uptake of the utterance which is responsive to its sequential implicativeness, we cannot classify the utterance as having been treated as non-serious (Holt, 2013b; Sacks, 1972; Schegloff, 2001). Instead, we can only say that it has been treated as a laughable, where not all laughables are treated as non-serious; indeed, there is an example of just such a sequence in Excerpt 7/8, below, in the talk prior to the utterance which is treated as non-

serious; I offer evidence to that effect as part of the analysis of that excerpt. Therefore, in conjunction with laughter, lack of serious response which is attentive to the (potentially) serious sequential implicativeness of the utterance is an important defining feature of the sequential context of non-seriousness.

In addition to laughter and lack of serious response, a final key characteristic of the sequential context of non-seriousness is a jointly managed return to seriousness. As a general rule, shifts of various sorts in conversation are rarely abrupt; for example, it has been well documented that shifts in topic generally occur in stepwise fashion (Sacks, 1987, 1995b), although certainly the conversational means exist to effect an abrupt shift in topic, albeit in marked fashion. As Schegloff writes, "the next thing to be done will bear the same accent as the prior did" (2001, p. 1954). Yet in the case of non-seriousness, for rather obvious interactional reasons, it is important to accomplish a clearly demarcated return-to-seriousness, that is, an ongoing non-serious status may be "defeasible, but it appears that it may need 'defeasing' [sic]" (Schegloff, 2001, p. 1954).

In the case of English, Schegloff (2001) has documented the use of a turn-initial "no" as one such means of "defeasing". In the data presented in that study, all instances of turn-initial negative particles marking a return-to-seriousness were deployed by the speaker of the utterance treated as non-serious, who then also made the first serious utterance. However, he notes that this action is not necessarily only done by the speaker of the non-serious utterance (p. 1948), and furthermore, that this is only one of perhaps many ways in which the return-to-seriousness may be jointly managed and accomplished in English (p. 1954). The key observation is that participants display an orientation to such a transition, which may or may not be accompanied by the deployment of a turn-initial "no", which in turn cannot be analyzed as signaling disagreement

with the prior utterance.

As I have argued elsewhere (Trygstad, 2014), and building on Betz' (2017) broader analysis of turn initial "*ja*" as a discourse marker, a similar model can be applied to German conversation, albeit with slight modification and expansion. In German, when a turn-initial particle is deployed in this sequential environment, the default, unmarked variant is the affirmative particle "*ja*" rather than the negative particle "*nein*". The negative particle is typically deployed in sequential contexts where an affirmative particle has the potential to be misconstrued as doing actual agreement (an example of which can be seen in Excerpt 7/8, below).

Where Schegloff (2001) limits his discussion to turn-initial "no" deployed by the speaker of the utterance treated as non-serious, who then also continues with the first serious utterance, I have proposed an expansion of his model to include turn-initial (as well as TCU-initial or "effective turn-initial" (Heritage, 1998) particles; see Excerpt 5/6, below, for an example of the latter) deployed by any participant in the conversation, or even by multiple participants (Trygstad, 2014). Furthermore, the speaker of the non-serious utterance and/or of the turn-initial particle need not necessarily be the one to continue talk with a serious utterance. An example of a sequence where both speakers in the dyad utter a turn-initial "*ja*", which is then followed by serious talk from the recipient of the original non-serious utterance rather than the speaker, can be seen in Excerpt 3/4, below.

The transition back to seriousness need not necessarily be marked with a turn-initial particle; it can also be accomplished via other means. In Excerpt 1/2, for example, we will see a return to seriousness done via an abrupt shift in topic after a gap of silence, with no turn-initial particle deployed by either participant. What is clear, however, is that there is a return-to-

seriousness which is jointly managed by the participants in that conversation. Such negotiation is the final hallmark which can be observed to signal that the prior utterance or spate of talk was treated as non-serious by the participants.

To summarize, the sequential features which must be present in order to permit the analysis that an utterance has been treated as non-serious by both participants in a dyadic conversation are as follows: both participants laugh (which is not treated as a source of trouble), there is no serious response to the (potentially serious) import and/or action of the utterance (which is not treated as a source of trouble), and there is a clear, jointly managed return-to-seriousness. In the sequential analyses presented below, each of these excerpts contains an utterance treated as non-serious which fits all of these criteria. As such, they are similar in kind and can be used as a basis for comparison of the acoustic features of non-serious utterances, as presented in the latter half of this chapter.

3.2.1. Sequential Analyses

In the following section, I present a sequential analysis of each utterance treated as non-serious, demonstrating that each fits the sequential criteria defined above. For each utterance treated as non-serious, there is a corresponding serious utterance from the same speaker selected from the immediately preceding talk for use as an experimental control in the acoustic analysis and perceptual experiment, as described in Chapter 2. In other words, there are a total of eight excerpts, of which four are treated as non-serious and four as serious. Each pair of utterances is from the same speaker, for a total of four unique speakers (two male, two female). All speakers are of a similar age and socioeconomic status (currently in post-secondary school, or recent graduates of a post-secondary course of study), are native speakers of German. Two regional colloquial varieties are represented, namely from the Berlin-Brandenburgish and the Hessian

areas, both standard varieties of Central High German, with one male and one female speaker of each.

The excerpts are numbered in such a manner as to identify first the non-serious excerpts, and then the serious excerpts which were used as the controls in the acoustic and perceptual experiments, such that the portions of each selection culled for use as the non-serious excerpts in the survey are identified with the odd numbers (1, 3, 5, 7), while the serious excerpts are identified with the corresponding even numbers (2, 4, 6, 8). Note that as a result, the evenly-numbered excerpts precede the odd-numbered excerpts in each selection. This is simply an artifact of the selection process, during which the non-serious utterances were selected first, and then a serious utterance from the same speaker was selected from the talk preceding the non-serious utterances, as described in Chapter 2. The excerpts used in the survey, i.e., the portions of the data which were made available to survey participants, are demarcated in the transcripts by indicating where the videos were trimmed with the notation ((Begin Excerpt X)) and ((End Excerpt X)). The precise non-serious and serious utterances which were trimmed and subjected to further acoustic analysis are indicated in the transcripts with an arrow (→). A full key of transcription symbols can be found in Appendix 3.1. A short description of the preceding conversation is given for each excerpt in order to provide additional context.

3.2.1.1. Excerpt 1/2: Eva und Simone 1 (21:12 - 22:16)

Eva and Simone are exchanging tellings about events in their respective childhoods. Simone has just completed a story in which she shares that she once filled a box with mussels and kept them in a cupboard without cleaning them first, which then subsequently began to smell very badly. After assessing this as "disgusting" ("*ekelhaft*"), Eva begins a new telling.

```
001  E    °·h:::° HUH::=wir haben früher auch immer    am
          °·h:::° HUH::=we  have earlier also always on the
          °·h:::° HUH::=we  always used to collect a ton of
```

002 ↑strand richtig viele ↑muscheln gesammelt, und
↑beach right many ↑mussels collected, and
↑shells at the ↑beach too, and

003 ↓dann wusste ich nicht was ich mit den ↑machen
↓then knew i not what i with them ↑do
↓then i didn't know what to ↑do with

004 sollte,
should,
them,

005 S [aja:¿]
[ahyeah:¿]
[ahyeah:¿]
[ ]
006 E [aber ich ↑hab s]ie gesammelt¿
[but i ↑have t]hem collected¿
[but i ↑colle]cted them¿

((Begin Excerpt 2))

007 S ich hab auch immer ↑stEIne gesammelt,
i have also always ↑stOnes collected,
i always collected ↑rOcks too,

008 [aber ich weiß nich-] ↑weiß nicht jetzt wo die
[but i know no-] ↑know not now where they
[but i don't kno-] ↑don't know now where they
[ ]
009 E [ja ich auch,]
[yeah i also,]
[yeah me too,]

010 S alle hin so[llen,]
all to shou[ld,]
all should [go,]
[ ]
011 E [immer] ganz viele weiße steine,
[always] quite many white stones,
[always] a whole bunch of white rocks,

012 (0.8)

013 S °ach nee° nicht nur ↑weiße die ↑hÜbschen
°oh no° not only ↑white the ↑prEtty
°oh no° not just the ↑white ones the ↑prEtty

→ **014** steine, ↑ich hab immer versucht ne
stones, ↑i have always tried a
rocks, ↑i was always trying

→ **015** verstEInerung zu finden,
fOssil to find,
to find a fOssil,

((End Excerpt 2))

016 (0.8)

017 **S** never found [one,]
never found [one,]
never found [one,]
[ ]
018 **E** [ich auch,]
[i also,]
[me too,]

019 (0.8)

020 **S** °·h:°
°·h:°
°·h:°

021 **E** ·ptch: ich hab immer versucht nen bernstein
·ptch: i have always tried an amber stone
·ptch: i was always trying

022 zu finden;
to find;
to find amber;

023 **S** °h:° never found one;
°h:° never found one;
°h:° never found one;

024 **E** muhuh.
muhuh.
muhuh.

025 (0.4)

026 **E** °·h::°
°·h::°
°·h::°

027 S ptch:[·h:]
ptch:[·h:]
ptch:[·h:]
[ ]
((Begin Excerpt 1))
[ ]
028 E [uh:m:] (1.5) ptch (0.5) ich hatte mal so
[uh:m:] (1.5) ptch (0.5) i had once so
[uh:m:] (1.5) ptch (0.5) i used to have

029 nen ↑richtig ↑coolen stein, °·ptch° der war so
a ↑right ↑cool stone, °·ptch° he was so
a ↑really ↑cool rock, °·ptch° it was so

030 rund¿ [°·ptch°] nicht s- nicht so ↑richtig
round¿ [°·ptch°] not s- not so ↑right
round¿ [°·ptch°] not s- not so ↑round
[ ]
031 S [mhmm,]
[mhmm,]
[mhmm,]

032 E ↑rund, aber halt so groß dass er in meine hand
↑round, but just so big that he in my hand
↑really, but so big that it fit in my hand

033 gepfasst hat so. genau¿
passed had so. exactly¿
just exactly. like so¿

034 S mhmm;=
mhmm;=
mhmm;=

035 E =wie ich so gemacht hab¿
=how i so did have¿
= just like this¿

036 S eheh
eheh
eheh

037 (0.3)

038 E und der hat ↑so: ↑schön ge↑glitzert?
and he has ↑so: ↑pretty spar↑kled?
and it spar↑kled ↑so: ↑beautifully?

039 S ·h:::,
·h:::,
·h:::,

040 E °·ptch° der war so ein bisschen ↑dUnkel↑blAU, und
°·ptch° he was so a little ↑dArk ↑blUE, and
°·ptch° it was like sort of ↑dArk ↑blUE, and

041 hat ↑richtig ↑heftig geglitzert, und ich wusste
has ↑right ↑heavy sparkled, and i knew
it sparkled ↑so ↑brilliantly, and i just

042 nur nicht (0.3) ich so was ist das für
only not (0.3) i so what is this for
didn't know (0.3) i was like what kind of

043 eigentlich ein stein. °·h::° hab ich ↑immer mit
actually a stone. °·h::° have i ↑always with
rock is this even. °·h::° i ↑always carried it

044 mir rumgetragen;
me around carried;
around with me;

045 (0.5)

→ **046 S** es war bestimmt en dia↑mAnte drinne, °( )° ☺
it was surely a dia↑mOnd therein, °( )° ☺
there was definitely a dia↑mOnd in it, °( )° ☺

((End Excerpt 1))

047 E jahuh sohuh never want to lose you, ☺ und jetzt
yeahuh sohuh never want to lose you, ☺ and now
yeahuh sohuh never want to lose you, ☺ and now

048 nur so miss him
only so miss him
just like miss him

049 S huhuhuhuh huh
huhuhuhuh huh
huhuhuhuh huh

050 E °·ptch:°=
°·ptch:°=
°·ptch:°=

051 S =jahuh auch nur so,
=yeahuh also only so,
=yeahuh just like that,

052 (1.1)

053 E ↑where's ↑eva, nksh::
↑where's ↑eva, nksh::
↑where's ↑eva, nksh::

054 (0.9)

055 E i miss her. [genau so, genau so,]
i miss her. [exactly so, exactly so,]
i miss her. [just like that, just like that,]
[ ]
056 S [da war]
[there was]
[there was]

057 bestimmt außerirdisches ↑leben drin, ☺
surely otherwordly ↑life therein, ☺
definitely alien ↑life in there, ☺

058 (0.5)

059 E bestimmt¿
surely¿
definitely¿

060 (0.7)

061 E oh gott,
oh god,
oh god,

062 (1.3)

063 S tja:, das ist möglich, ☺
tyeah:, that is possible, ☺
tyeah:, it's possible, ☺

064 (0.5)

065 E ptch: (0.6) sag mal, (0.8) we:rden wi:r (0.1)
ptch: (0.6) say once, (0.8) are: we: (0.1)
ptch: (0.6) remind me, (0.8) are: we: (0.1)

066 gefi:lmt oder wird damit-= >was wird
fi:lmed or is therewith-=>what is
being fi:lmed or is it-= >what is being

067 damit gemacht?<
therewith done?<
done with it?<

After a quiet inhalation and relatively loud sigh, Eva begins her telling by introducing a related topic, namely, that she also used to collect a lot of shells at the beach and did not know what to do with them (lines 1-5, 6). Simone offers a stepwise transition to a related topic (Sacks, 1987, 1995b), indicating that she collected stones, and also did not know what to do with them (7-8, 9). In this way, Eva and Simone display alignment by engaging in parallel tellings of things which they collected as children, despite such collecting making relatively little sense in retrospect.

In line 11, Eva specifies that she preferred to collect stones which were white in color ("*weiße steine*"). After a gap of 0.8 seconds, indicating that a dispreferred action is potentially forthcoming (Pomerantz, 1984; Auer & Uhmann, 1982), Simone states that she not only collected white stones, but "pretty" ones ("*hübschen*"; line 13) as well, expressing mild disagreement by broadening the category of desirable stones previously established by Eva. Simone then adds to this statement that she always looked for a "fossil" ("*versteinerung*"; 14), offering her reason for collecting stones while also again effectively narrowing the category of stones sought and expressing disagreement with Eva that the most desirable stones were only white ones. Subsequent to a gap of 0.8 seconds, again, indicating that there may be a dispreferred action forthcoming, Simone self-selects as next speaker and, in the first instance of code-switching to English which is characteristic of this excerpt, indicates that she never found a fossil (17). In overlap, Eva states that she also looked for fossils, and after a gap, indicates that she looked for "amber" ("*bernstein*"; 21) as well. Simone supplies the potential result that Eva

also never found amber, which Eva receipts positively in line 24 ("*muhuh*"), indicating that she did not, in fact, ever find amber, and effectively bringing the sequence to a close. In this way, Eva treats the action and import of Simone's utterances concerning her search for a fossil as serious by contributing to the ongoing telling and indicating the types of stones that she sought apart from white ones, i.e., fossils and amber.

In line 28, Eva again initiates a longer telling, making a bid for the floor with a protracted "*uhm*". She then begins to describe a specific stone that she had previously had. The stone is described positively as "really cool" ("*richtig coolen*"; 29) and having "sparkled so beautifully" ("*so schön geglitzert*"; 38). Throughout this description of the stone, Simone produces a series of continuers (lines 31, 34, 36, and 39), thereby continuing the suspension of the turn-taking procedure in order to allow the telling to proceed (Mandelbaum, 2012). Additionally, Simone's extended inhalation in line 39, through the slight rise in intonation, shows positive alignment with Eva's positive assessment of the stone in line 38.

Eva continues her description of the stone, stating that it was "dark blue" in color ("*dunkelblau*"; 40) and providing an upgrade of her previous assessments (Pomerantz, 1984) by saying that the stone "had sparkled so brilliantly" ("*hat richtig heftig geglitzert*"; 41). She then brings her telling to a close by indicating that she did not know what sort of stone it was (41-43), abandoning the original telling format ("*ich wusste nur nicht*", "I just didn't know"; 41-42) in favor of an enactment of her reaction to the stone. This is done by embedding a grammatical question in the ongoing talk in lines 42-43 in a reported speech format: "I was like, what kind of rock is this even" ("*ich so was ist das für eigentlich ein stein*").

After a gap of 0.5 seconds, Simone provides a candidate response to Eva's embedded question by stating that "there was definitely a diamond in it" ("*es war bestimmt en diamante*

drin"; 46), her utterance then trailing off in non-speech tokens. This utterance can be heard to differ prosodically from the surrounding talk, and has been marked as a potential instance of "smile voice" with the ☺ symbol. Indeed, Eva treats her utterance as non-serious. Note that she does not respond to the potentially serious action that Simone's utterance could have done, namely, providing a candidate response to the embedded question of what sort of stone this was. Instead, rather than agreeing or disagreeing with Simone's suggestion that the stone could have had diamonds, Eva first laughs (47) and engages in a sort of role-play where she speaks to the stone in English (lines 47-48), saying that she didn't ever want to lose the stone, but now only misses it (presumably because it was lost eventually); this sort of hypothetical discourse is relatively common in joke situations (Golato, 2012). Eva's laughter and subsequent talk in essence "detopicalize" the question of what sort of stone this was, making irrelevant any serious uptake of Simone's candidate response and thereby treating it as non-serious (Sacks, 1972; Schegloff, 1987).

Simone, for her part, treats Eva's role-play with the stone as non-serious by laughing in response (lines 49 and 51). Eva continues the role-play by speaking now from the stone's perspective in lines 53 and 55, indicating the stone cannot find her and misses her as well. Simone then provides an additional candidate response to the question of what sort of stone this was by stating that there was certainly "alien life" ("*außerirdisches leben*"; 57) in the stone, which is likely responsive to Eva's depiction of the stone speaking. Again, this is uttered in a manner which differs prosodically from the surrounding talk, and has been marked as a potential instance of "smile voice". Perhaps because this suggestion is substantially more unlikely than the suggestion that a stone may contain diamonds, Eva this time does not respond with laughter, but rather confirms this interpretation (59) followed by the phrase "oh god" ("*oh gott*"; 61). In

this way, she does not overtly display her orientation to Simone's utterance in line 57 as non-serious, for example, by laughing, but rather seems to be continuing to engage in non-seriousness through a mock display of agreement and fear, which in turn underscores the absurdity of the conversation itself.

Finally, in line 63, Simone brings the sequence to a close by uttering the token "*tja*" (a discourse marker indicating changes in topic and/or action sequence; Zifonun, Hoffmann, & Strecker, 1997) and stating that it is possible ("*das ist möglich*"). The placement of the turn initial discourse marker "*tja*" assists in managing what Schegloff (2001) has described as the transition from non-serious talk back to serious talk, although she does continue with a non-serious utterance in this turn, which is a slightly unusual move (potentially due to the use of "*tja*" as opposed to "*ja*" in this sequential environment). Despite this, the transition is in fact successfully accomplished, albeit after a gap of silence (line 64), in that Eva proceeds in the next turn at talk with a question about the future fate of the data being recorded (65-67). Talk continues at this point with serious discussion of the nature of the study they are participating in and speculation about what will happen with the data.[1]

On the basis of this analysis, we can conclude that Simone's utterance in line 46 is treated as non-serious by both participants in the conversation, and is suitable for further acoustic analysis on the basis of the selection criteria outlined above. Conversely, we observed that her utterance in lines 14-15 was treated as a serious action, and therefore can be considered suitable for use as a control value to compare with the acoustic features of Simone's non-serious utterance.

[1] Measures were taken subsequent to this recording in order to ensure that the participants knew the answers to these questions and to reestablish informed consent.

3.2.1.2. Excerpt 3/4: Maximilian und Emanuel 1 (04:37 - 05:35)

Maximilian and Emanuel are discussing the play strategies utilized by various European soccer teams. Specifically, they are debating the merits of a strategy commonly associated with the Spanish teams, namely, the *tiki-taka*. This strategy consists of passing the ball and maintaining possession until an opportunity to score a goal presents itself opportunistically, obviating the necessity of using a striker to score a goal by actively penetrating the defenses of the opposing team (Ball, 2014; Lowe, 2010).

001 **E** um::. (0.5) (ja ich mein gut) man kann sagen
um::. (0.5) (yeah i mean good) one can say
um::. (0.5) (yeah i mean well) you could say

002 >ne< bei den ↑spaniern ↑klappts ja,
>nah< by the ↑spaniards ↑works it yeah,
>nah< it ↑works for the ↑spaniards huh,

003 **M** nja. das- die sind nicht schlecht damit gefahren¿
myeah. that- they are not bad therewith gone¿
well. that- they haven't done poorly with it¿

004 (0.3)

005 **M** ich weiß- ich ha- ich hab en [tierischen]
i know- i ha- i have a [bestial]
i know- i ha- i have a [brutal]
[ ]
006 **E** [abe:r >ganz-<]
[bu:t >quite-<]
[bu:t >quite-<]

007 ehrlich >wenn ich mir so barcelona ↑anguck,<
honestly >when i me so barcelona ↑watch, <
honestly >when i watch like ↑barcelona, <

008 (0.1) >>ab und zu mal < könnt ich mir
(0.1) >>off and on once< could i me
(0.1) >> sometimes < i could

009 ↑vorstellen:,<= dann nimm doch wenigstens <einen
↑imagine:, <= then take yet at least < one
↑imagine:, <= then just take at least < one

010 mit >, (0.3) und wenns nicht läuft, dann
with>, (0.3) and if it not runs, then
with>, (0.3) and it if doens't work, then

011 [wechselt] den ein, stellt den vorne >↑hin < und
[switch] him in, set him front >↑to < and
[switch] him in, put him out up >↑front< and
[ ]
012 M [mhmm]
[mhmm]
[mhmm]

013 E spielt dem die dinger auf die birne,
play him the things on the pear,
pass it to him for a header,

014 [(dass de- die reinmacht)]
[(that th- they in do)]
[(that th- they rush in)]
[ ]
((Begin Excerpt 4))
[ ]
015 M [die sache ist eigentlich] (0.1) ich kA:nns
[the thing is actually] (0.1) i cA:n it
[the thing is actually] (0.1) i cA:n't

016 nicht mehr sehen, ich habs mir an der ee em.=
not more see, i have it me at the ee em.=
watch it anymore, i watched it during the EM. =

017 =also, °·h::°=
=also, °·h::°=
=also, °·h::°=

018 E =das tiki-taka¿=
=the tiki-taka¿=
=the tiki-taka¿=

019 M =das tiki-taka; also- ich- ich war auch uff da
=the tiki-taka; so- i- i was also uff there
=das tiki-taka; well- i- i was also uff there

→ **020** uff tOll wie die den bAll laufen lassen, und
uff grEAT how they the bAll run let, and
uff grEAT how they run the bAll around, and

→ **021** dann ham die es (0.1) wirklich so bis zum
then have they it (0.1) really so up to
then they did it (0.1) really to the point of

→ **022** erbrE:chen gemacht bei der ee em [irgend]wie,
vomI:ting done by the ee em [some]how,
vO:miting at the EM s[omethi]ng like,
[ ]
023 **E** [ja,]
[yeah,]
[yeah,]

→ **024** **M** achtzig prozent bA:llbesitz, und (0.3) zwEI:
eighty percent bA:ll possession, and (0.3) twO:
eighty percent bA:ll possession, and (0.3) twO:

025 torschüsse, ·h:::
goal shots, ·h:::
shots at the goal, ·h:::

((End Excerpt 4))

026 **E** ja. =
yeah.=
yeah.=

027 **M** =und irgendwie=ja, und dann wird nOchmAl
=and somehow =yeah, and then is agAIn
=and somehow =yeah, and then agAIn there's more

028 hinten ↑rumge↑spielt und dann jetz=schIEß
behind ↑around ↑played and then now =shOOt
↑playing ↑around backfield and then like= just

029 doch einer mal. bitte,
yet one once. please,
shOOt one now. please,

030 **E** huh
huh
huh

031 **M** schIEß doch EInfach mAl den bAll, ich wIll nicht
shOOt yet SImply Once the bAll, i wAnt not
just shOOt the bAll One tIme, i cAn't take it

032 mehr,° ·h::°
more,° ·h::°
anymore,° ·h::°

033 **E** ja?h:,
yea?h:,
yea?h:,

034 (0.6)

035 **E** huhuhuh huhuhuh [·h::]
huhuhuh huhuhuh [·h::]
huhuhuh huhuhuh [·h::]
[ ]

((Begin Excerpt 3))

[ ]
036 **M** [ja] und jetzt >nimm doch auch
[yeah] and now >take yet also
[yeah] and now >just take out

037 noch < den einzelnen- den einzigen stürmer raus,
still< the single- the only striker out,
the < single- the only striker too,

→ **038** alles klar; [ja:] noch- noch mehr: bAll hin
all clear; [yea:h] still- still more: bAll to
sounds good; [yea:h] even- even more: bAll back
[ ]
039 **E** [huhuh]
[huhuh]
[huhuh]

→ **040** **M** und herschieben, ↑dAs hat uns gefehlt; h::: ☺
and fro pushing, ↑thAt has us lacked; h::: ☺
and forth, ↑thAt's what we were missing; h::: ☺

((End Excerpt 3))

041 **E** huh huhuhuh ·h:: ja?h,
huh huhuhuh ·h:: yea?h,
huh huhuhuh ·h:: yea?h,

042 (0.2)

043 **M** ja.
yeah.
yeah.

044 (1.4)

045 **E** aber uhm wen ich ↑krAss fand war ja ↑drAxler¿
but uhm whom i ↑crAss found was yeah ↑drAxler¿
but uhm who i thought was ↑tErrific was ↑drAxler¿

046 (1.2)

047 der ist zUr zEIt ja Echt gUt drauf,
he is At tIme yeah rEAlly gOOd thereon,
he is doing rEAlly gOOd right nOw,

The excerpt opens with a positive assessment from Emanuel, stating that the tiki-taka strategy works well "for the Spaniards" ("*bei den spaniern*"; line 2). Maximilian responds with a slightly downgraded agreement prefaced with the token "*naja*", a strong indicator of a forthcoming downgrade (Auer & Uhmann, 1982; Golato, 2018), indicating that "they haven't done poorly with it" ("*die sind nicht schlecht damit gefahren*"; 3). While this second assessment is in agreement with Emanuel's assessment, it is downgraded in that it implies that, although they have not done poorly, Maximilian still cannot assess it as especially good. After a gap of 0.3 seconds, Maximilian begins a new utterance (line 5). This utterance is not brought to completion, however, as Maximilian cuts off and self-repairs "*ich weiß*" ("I know") to "*ich hab*" ("I have"), at which point Emanuel interrupts him (6). Here, he describes how the Barcelona team could use the tiki-taka strategy to send a striker forward and score a goal (6-11, 13-14). This utterance, given the sequential position subsequent to the downgrade of Emanuel's original assessment, serves as a means of reestablishing alignment with Maximilian by offering an indirect criticism of the tiki-taka strategy via a suggestion to the Barcelona team to make use of a striker.

In overlap with the end of Emanuel's turn at talk, Maximilian again issues an assessment of the soccer strategy. He offers here a stronger negative assessment by displaying disgust with the strategy, stating that he "can't watch it anymore" ("*kanns nicht mehr sehen*"; 15-16). In line

18, Emanuel says "the tiki-taka" ("*das tiki-taka*") with rising intonation, which Maximilian then repeats with slightly falling intonation (19), serving to reaffirm topical alignment.

Maximilian then provides a longer explanation of how he came to have such a negative opinion of the strategy, beginning in line 19. He first indicates that he originally thought it was "great" ("*toll*"; 20) how the Spanish would "run the ball" ("*den ball laufen lassen*"; 20), issuing for the first time a positive assessment of the strategy in question. Subsequent to this positive assessment, Maximilian then does a short telling in which he recounts how the Spanish team used the tactic during the recent Union of European Football Associations European Championship of 2012, when the Spanish national team won (UEFA Euro Cup, 2012). He states that they used the strategy "to the point of vomiting" ("*bis zum erbrechen*"; 21), a clearly negative assessment of the strategy. This assessment is based on the assertion that they had possession of the ball 80% of the time but only made two shots at the goal (lines 22 and 24), which, from the perspective of a spectator, provides the grounds for Maximilian's negative assessment of the strategy as boring to watch, regardless of its potential effectiveness or role in helping Spain to win the championship. Maximilian continues the telling by describing how the Spanish national team continued "playing around" ("*rumgespielt*"; 28) at the back of the field, that is, away from the goal and out of range of the possibility to score. He then provides an account of hypothetical discourse (Golato, 2012), in which he implores the team to shoot the ball (28-29, 31), and indicates he "can't take it anymore" ("*will nicht mehr*"; 31-32), all of which serve as further negative assessments of the tiki-taka strategy as boring to watch.

Throughout this extended telling, Emanuel utters several acknowledgment tokens at lines 23, 26, and 33. These serve to acknowledge receipt of the information which Maximilian is conveying and also as continuers which support continued suspension of turn-taking for

Maximilian's extended telling. Through line 33, both participants treat the conversation as serious, insofar as neither of them treat any utterance as a laughable by responding with laughter.

Following a 0.6 second gap, Emanuel laughs for the first time (35), encouraging Maximilian to further act out his hypothetical discourse (Golato, 2012) while also signaling the transition to Emanuel's treatment of the utterances as non-serious. Maximilian continues his telling about the European Championship by stating that the team should take the only striker out of the game (36-38), which he assesses as positive with "sounds good" ("*alles klar*", 38). This mock suggestion and assessment is made despite his previous assertions that the lack of a striker is not favorable, and can therefore be understood as sarcastic. Finally, in lines 38 and 40, using a slightly differing prosody from the surrounding talk, Maximilian utters the line of interest to the current analysis: "yeah, even more ball back and forth, that's what we were missing" ("*yeah noch noch mehr ball hin und herschieben das hat uns gefehlt*").

This utterance is of particular interest because it is overtly non-serious when viewed in the context of the preceding telling. Throughout the sequence, Maximilian has expressed frustration that the Spanish national team had moved the ball around too much without taking shots at the goal. However, he indicates here that moving the ball around was precisely what was missing in the game. This utterance can therefore be classified as non-serious in that it conflicts strongly with the assertions that Maximilian previously made that there was too much ball movement.

Indeed, we observe that both participants treat the utterance as non-serious, confirming the above analysis. In line 41, Emanuel laughs and issues the token "*ja*" ("yeah"). As described previously, by laughing, he treats the preceding utterance as non-serious, eliminating the relevance of a serious response on his part to the action of the preceding turn (Sacks, 1972;

Schegloff, 1987). Maximilian then also utters the token "*ja*" in line 43. By doing so, he also treats his previous utterance as non-serious (by receipting Emanuel's laughter as having been the relevant next response), and transitions back to serious talk via a sequence closing third.

Following a longer gap of 1.4 seconds, Emanuel offers a positive assessment of a player on the German national team, Julian Draxler (45). He indicates that he is "doing really good right now" ("*zur zeit echt gut drauf*"; 47). Talk then continues with Maximilian agreeing and discussion of what Draxler is currently doing well. In this excerpt then, we see that the speakers engage in a jointly managed transition back to seriousness. The "*ja*" tokens in lines 41 and 43 served to manage the transition from non-seriousness back to seriousness, while also closing the previous sequence. Talk then continues after selection of a new topic (i.e., assessment of a player on a different team).

In the preceding excerpt, we therefore conclude that lines 38 and 40 contain an instance of non-seriousness, as judged by how the speakers treat the utterance. The utterance in lines 19-22 and 24, on the other hand, is treated as a serious assessment by both speakers. These excerpts therefore fit the selection criteria established in Chapter 2 and can be measured acoustically and utilized in the perceptual survey, as described below.

3.2.1.3. Excerpt 5/6: Donnie und Gerhart 1 (07:12 - 08:02)

Donnie and Gerhart are engaged in conversation about a topic wholly unrelated to the excerpt, specifically the nature and status of homophobia in Germany today. At a lull in the conversation, Donnie indicates that he is changing the topic of conversation entirely in order to discuss the forthcoming events of the evening.

```
001  D    mal  ↑thEmawechsel  weil  ich da   oben grad den
          once ↑tOpic change because i there over now  the
          quick ↑tOpic change because i'm looking up there
```

002 alkohol sehe, (0.4) haben huh wir huh für huh
alcohol see, (0.4) have huh we huh for huh
at the alcohol, (0.4) do we huh even huh have huh

003 heut huh abend huh überhaupt huh genug huh
today huh evening huh at all huh enough huh
enough huh alcohol huh for huh this huh

004 alkohol huh huh¿ [h:huhuhuhuh]
alcohol huh huh¿ [h:huhuhuhuh]
evening huh huh¿ [h:huhuhuhuh]
[ ]
005 G [h:huhuhuhuh]
[h:huhuhuhuh]
[h:huhuhuhuh]

006 ·h:: >>wir haben<< nur noch die: [um::] (0.3)
·h:: >>we have << only still the: [um::] (0.3)
·h:: >>we only << have left the: [um::] (0.3)
[ ]
007 D [·h::] (d)
[·h::] (d)
[·h::] (d)

008 G die vodka=>↑da ist aber nicht mehr viel
the vodka=>↑there is but not more a lot
the vodka=>↑but there isn't very much left

009 drin, < [also es reicht dann.]
therein,< [so it reaches then.]
there, < [well but it's enough.]
[ ]
010 D [ick wollt grad sagen] das ist doch
[i wanted now to say] that is yet
[i was just about to say] but that's

011 auch nur noch hier son bisschen
also only still here so a little bit
also only really a little bit here

012 G >aber er hat schon recht dass dass< wir nicht
>but he has already right that that< we not
> but actually he's right that that < we

013 zu vie:l uh verschenken dürfen,
too mu:ch uh give away may,
shouldn't uh give away too mu:ch,

014 (0.7)

015 **G** wenn hannelie kommt¿
if hannelie comes¿
if hannelie comes¿

016 (0.4)

017 **G** °der hoschi weggeht ja¿°
°the buddy goes away yeah¿°
°the buddy goes away yeah¿°

018 **D** ʔhuh
ʔhuh
ʔhuh

019 (1.5)

020 **G** >ich weiß jetzt n- nicht was wir machen sollte-<=
> i know now n- not what we to do shoul-<=
> i don- don't know now what we should do- <=

((Begin Excerpt 6))

021 **G** ↑ich weiß nicht, ich kann ihn nicht einschätzen;
↑ i know not, i can him not assess;
↑i don't know, i can't figure him out;

022 **D** ich auch nich.
i also not.
me either.

023 (0.4)

→ **024** **G** [er ist] wie ↑ich ohne zu reden;
[he is] how ↑i without to talk;
[he is] like ↑me without talking;
[ ]
025 **D** [(also)]
[(so)]
[(well)]

((End Excerpt 6))

026 (0.3)

027 **D** °j-° (.2) ja [uh:] er-(0.2) grI:nst ↑immer so:,
°y-° (.2) yeah [uh:] he-(0.2) grI:ns ↑always so:,
°y-° (.2) yeah [uh:] he-(0.2) ↑always grI:ns so:,
[ ]
028 **G** [huh]
[huh]
[huh]

029 (0.3)

030 **D** grenzdebil¿ hu[huh]
borderline idiotic¿ hu[huh]
borderline idiot¿ hu[huh]
[ ]
031 **G** [grenz]debil; ja,
[borde]rline idiotic; ya,
[borde]rline idiot; ya,

032 (0.1)

033 **G** oder schon debil. huhuh
or already idiotic. huhuh
or just an idiot. huhuh

034 (1.0)

035 **G** se:nIl, huhuh=
se:nIle, huhuh=
se:nIle, huhuh=

036 **D** =tsuhtsuh °zum glück ist dat fenster
=tsuhtsuh °to luck is the window
=tsuhtsuh ° luckily the window is

037 anjeklappt¿°=
shut¿° =
shut¿° =

038 **G** = ja,
=yeah,
=yeah,

039 (0.3)

040 **D** th huhuhuh
th huhuhuh
th huhuhuh

041 (0.2)

042 **G** er hat sich auch n- > ↑nie wirklich bei uns-<
he has himself also n- >↑never really by us-<
he also has himself n- >↑never really to us-<

043 (0.1) ↓er hat sich vorgestellt.=ok,=gut. aber,
(0.1) ↓he has himself introduced. =ok,=good. but,
(0.1) ↓he has introduced himself. =ok,=good. but,

044 (0.4)

((Begin Excerpt 5))

045 **G** ich weiß so gut ↑wie gar nichts, hast du
i know so good ↑how absolutely nothing, have you
i know as good ↑as nothing, have you added

046 ihn bei facebook mal geaddet¿ irgendwie [(so)?]
him by facebook once added¿ somehow [(so)?]
him on facebook yet¿ something like th[(at)?]
[ ]
047 **D** [nee.]
[nah.]
[nah.]

048 (0.3)

049 **D** ↑also ick weiß dass er informAtik studiert,
↑so i know that he computer scIence studies,
↑well i know that he's studying computer scIence,

050 oder irgendwie so was¿
or somehow so what¿
or something like that¿

051 (0.4)

→ **052** **G** wie uncool, ☺
how uncool, ☺
how uncool, ☺

053 (0.2)

→ **054** **G** wer macht denn ↑dAs bitte. ☺
who does then ↑thAt please. ☺
who does something like ↑thAt. ☺

((End Excerpt 5))

```
055        (0.6)

056  D     ↑uhwewe.=
           ↑uhwewe.=
           ↑uhwewe.=

057  G     =uh huhu[huh]
           =uh huhu[huh]
           =uh huhu[huh]
                   [   ]
058  D             [huh]uh[uh]
                   [huh]uh[uh]
                   [huh]uh[uh]
                          [  ]
059  G                    [do]nnie:, ·h:: ok  ja,
                          [do]nnie:, ·h:: ok yeah,
                          [do]nnie:, ·h:: ok yeah,

060  D     °·h:°
           °·h:°
           °·h:°

061  G     und  dann¿
           and  then¿
           what else¿

062        (0.7)

063  D     °h:° ↑joa, das  weeß ick nick dann.
           °h:° ↑yah, that know  i  not  then.
           °h:° ↑yah,   then i don't know.

064  G     er- kommt aus ↑stAH:nsdorf? oder;
           he- comes out ↑stAH:nsdorf?  or;
           he-  is from  ↑stAH:nsdorf? right;

065  D     stahnsdorf.  ja,
           stahnsdorf. yeah,
           stahnsdorf. yeah,
```

The excerpt begins where Donnie overtly announces that he is changing the topic ("*themawechsel*"; line 1). He first provides an account for this shift by stating that it is because he sees the alcohol "up there" ("*da oben*"; 1), then after a pause of 0.4 seconds, he asks whether

the alcohol they have is "enough for this evening" (2-3). This utterance is infused with laugh tokens and followed by outright laughter from both Donnie and Gerhart. The status of this utterance as serious or non-serious is initially unclear for the outside analyst, but it is certainly treated as humorous to this point. Donnie and Gerhart then agree that there is not a lot of alcohol (vodka is specifically mentioned; 8), but that this is sufficient for that particular evening (6-11).

In lines 12-13, Gerhart then indicates that "he's right" ("*er hat schon recht*") that they cannot give away too much alcohol. This provides a further clue that Donnie's question in lines 2-3 was in fact serious, as we can infer that an unnamed non-present third party has potentially expressed doubts previously that the alcohol on hand would be sufficient for that evening. The reasons why this was nonetheless treated as humorous by the conversationalists, unfortunately, remain unclear. However, the question as to whether there is sufficient alcohol is potentially delicate in nature, such that the laugh tokens may indicate embarrassment and acknowledgment of the topic as delicate, while the initial question about the amount of alcohol present as being sufficient for the evening had serious import and was treated as such. The participants' conclusion then seems to be that the quantity of alcohol on hand is sufficient for the evening, assuming that they abide by the non-present third party's suggestion not to give too much away over the course of the evening.

After a gap of 0.7 seconds, Gerhart self-selects as next speaker and says, "if Hannelie comes" ("*wenn hannelie kommt*"; 15), followed by a gap and then, in a quieter voice, by the utterance "the buddy goes away" ("*der hoschi weggeht*"; 17). Gerhart here makes clear that it is anticipated that Hannelie will come and is welcome to share the limited alcohol supplies, but that an unnamed individual is welcome neither to the alcohol nor at the party. Note that Gerhart indicates that the buddy should go away of his own accord, not that he is not invited. In other

words, he will not be sent away if he chooses to come, but his presence is clearly presented as undesirable.

In response to this, Donnie produces a minimal laugh token (18), followed by a gap of 1.5 seconds. Gerhart then indicates that he does not know what they should do about "the idiot" (20), and that he "can't figure him out" ("*kann ihn nich einschätzen*"; 21), a perhaps ironic choice of words as this utterance initiates a jointly constructed negative assessment. Gerhart, who is known to his friends as extremely outgoing and verbose, first states that he is like him, but "without talking" ("*ohne zu reden*"; 24). This serves as a negative assessment, i.e., that he does not engage with or speak to them, which appears to be a central problem in the relationship Donnie and Gerhart have with him.

Although perhaps this was an attempted joke, Donnie ultimately treats Gerhart's initial negative assessment as serious by producing an acknowledgment token ("*ja*"; 27). This is followed by his own negative assessment that he "always grins" ("*grinst immer so*"; 27), which he builds up incrementally by assessing him as a "borderline idiot" ("*grenzdebil*"; 30), in the sense of grinning in a borderline idiotic way. Gerhart first receipts this negative assessment by repeating it (line 31) and then upgrades it with a qualified agreement (Auer & Uhmann, 1982) by saying that he might be "just an idiot" ("*schon debil*"; 33). After a gap of 1.0 second, Gerhart suggests that he might be "senile" ("*senil*"; 35) instead, followed by laugh tokens. Donnie then laughs, looks to the window (which opens onto a balcony shared with the entire floor), and quietly notes that it is closed, "luckily" ("*zum glück*"; 36). This entire spate of talk serves as an ongoing complaint about and negative assessment of the individual in question, on the basis of the claim that he is unintelligent.

Gerhart then produces an acknowledgment token (38) while Donnie again laughs (40),

after which Gerhart begins to make the complaint that the buddy has never really introduced himself to them (42), which is then self-repaired in line 43. Here, it is implied that perhaps this person has minimally introduced himself for example by name, but has not made an effort to build a meaningful relationship with Gerhart and Donnie, despite apparently being often present. Following a gap of 0.4 seconds, Gerhart indicates that he knows nearly nothing about him and inquires if Donnie has added him on Facebook (45-46). In this way, Gerhart has returned to the original negative assessment, that is, he does not know him well and has been unable establish any sort of relationship with him.

Donnie treats Gerhart's question about Facebook as a true information question by responding in the negative (47). After a gap of 0.3 seconds, Donnie then supplies some information about their acquaintance, but marks it as uncertain, indicating his belief that he studies "computer science or something like that" ("*informatik oder irgendwie so was*"; 49-50). Following a further gap of 0.4 seconds, Gerhart produces a negative assessment of this information, saying "how uncool / who does something like that" ("*wie uncool* / *wer macht den das bitte*"; 52 and 54). This negative assessment of the subject area of computer science as "uncool" is uttered in a voice which differs slightly from the surrounding talk, and has been marked as a potential example of smile voice.

Despite the audible variation in prosody, the negative assessment is followed by a gap of 0.6 seconds, indicating that a dispreferred response may be forthcoming. Donnie, himself a computer science major, then produces a series of nonsensical, mimicry-type syllables (56), which could potentially be interpreted by Gerhart as an indication that he is initially treating Gerhart's negative assessment of computer science as serious. At this point, Gerhart treats his own utterance as non-serious by laughing (57), effectively inviting Donnie to also treat the

utterance as non-serious and laugh in response (Jefferson, 1979), an invitation which he accepts by laughing in slight overlap with Gerhart's laughter (58). Although there was initially evidence of trouble (lines 55-56) in determining the serious / non-serious status of Gerhart's negative assessment, ultimately both parties treat it as non-serious by laughing. Furthermore, there is no serious uptake of Gerhart's turn (Sacks, 1972; Schegloff, 1987), such as a second assessment, as would be appropriate following an initial assessment (Pomerantz, 1984).

In addition to treating the utterance as non-serious by laughing and neglecting to produce any serious uptake of the assessment action, Gerhart also produces a "*ja*" token (59), followed by the question "what else" ("*und dann*"; 61). This serves as a means of transitioning from the non-serious utterance back to serious talk (Schegloff, 2001), which is potentially also assisted by the use of the address term Donnie in line 59 (cf. Betz, 2017). In this excerpt, however, this transition is not accompanied by a topic shift, as seen in the previous excerpts, but rather a return to topic. Following a gap of 0.7 seconds, Donnie indicates that he does not know anything more about their acquaintance, treating Gerhart's inquiry in line 61 as serious (indicating a successful transition back to seriousness) while also making a claim of no knowledge. Gerhart then supplies a candidate response to his own request for more information by offering the name of their acquaintance's hometown, followed by a particle requesting confirmation of this information, similar to the English "right" ("*oder*"; 64). Donnie confirms that this is in fact his hometown, and talk continues with further discussion about what information is known about the individual in question.

The above analysis demonstrates that the utterance of interest in lines 52 and 54 is in fact treated by the both participants as non-serious by producing laughter, despite some initial indications of a potential source of trouble. The utterance in line 24, on the other hand, is treated

as a serious conversational action, with further assessments produced in response to the initial assessment. Both utterances are therefore suitable for further consideration in the present study.

3.2.1.4. Excerpt 7/8: Elena und Marion 1 (18:34 - 19:56)

Elena and Marion are conversing about the apartment Marion will soon be moving into, where her new roommate, Brunhild, already lives. Marion has been describing the various items she had purchased that day at the hardware store to decorate and furnish her new apartment, and also how she intends to paint her new room with decorative pink stripes (an activity which was to take place the next day after this recording was made). Both speakers have already bemoaned the task of painting the room, but conclude that it could be worse, differentiating between awful ("*schlimmes*") painting requiring spackle and such, as opposed to the slightly less arduous task of painting decorative stripes.

```
001  M    °ayayay.°
          °ayayay.°
          °ayayay.°

002  E    och wie aufregend,
          ah  how exciting,
          ah  how exciting,

003  M    [ja::.            ]
          [ja::.            ]
          [ja::.            ]
          [                 ]
004  E    [> da   wohnt ihr ]  da  < beide in der spelunke,
          [>there live y'all] there< both  in the dive bar,
          [>you two will be ]living< there in  a  dive bar,

005  M    °in der spelun[ke°]
          °in the dive b[ar°]
          °in  a  dive b[ar°]
                        [   ]
006  E                 [huh]uhuhuhuhuhuhuhuhuhuh
                       [huh]uhuhuhuhuhuhuhuhuhuh
                       [huh]uhuhuhuhuhuhuhuhuhuh
```

007 °huhuhuh° [°·huh°]·huh·huh·huh[·huh°huhuhuh° ·h:]
°huhuhuh° [°·huh°]·huh·huh·huh[·huh°huhuhuh° ·h:]
°huhuhuh° [°·huh°]·huh·huh·huh[·huh°huhuhuh° ·h:]
008 M [·h::] [ ]
[·h::] [ohje huhuhhuhuh·h]
[·h::] [oiye huhuhhuhuh·h]
[·h::] [oiye huhuhhuhuh·h]

009 (0.4)

010 M °in der spel-° oh da könnt en spe↑lUnkenrat
°in the dive-° oh there could a dive↑ bAr advice
°in the dive-° oh there could be a dive↑ bAr

011 geben,
give,
advice group,

012 (0.4)

013 M °huhuhuhuh°
°huhuhuhuh°
°huhuhuhuh°

014 E m↑hmm¿
m↑hmm¿
m↑hmm¿

015 (1.1)

016 E und dann k- und dann schenken wir euch ↑auch so
and then k- and then give we y'all ↑also so
and then k- and then we'll give y'all ↑also like

017 en- so en aschenbecher,
a- so an ash tray,
a- like an ash tray,

018 (0.7)

019 E so en stammtischaschen[becher,]
so a regulars' table ash[tray,]
like a stammtisch ash[tray,]
[ ]
020 M [oh das]
[oh that]
[oh that]

021 E [und huh stellen huh wir huh]
[and huh place huh we huh]
[and huh we'll huh put huh]
[ ]
022 M [wäre toll]
[would be great]
[would be amazing]

023 E ihn huh aufn huh tisch, huh
it huh on the huh table, huh
it huh on the huh table, huh

((Begin Excerpt 8))

→ **024 M** ich hätt auch [gern] so en schI:ld für den tisch;
i have also [like] so a sI:gn for the table;
i'd also like [to h]ave a sI:gn for the table;
[ ]
025 E [·h::]
[·h::]
[·h::]

026 (0.3)

→ **027 M** weißte mit so nem stAmmtischschild, ·h:
know you with so a rEgulars' table sign, ·h:
y'know with like a stammtisch sign, ·h:

((End Excerpt 8))

028 E m[hmm,]
m[hmm,]
m[hmm,]
[ ]
029 M [mit so] en wappen das wir aufhuh[häng]enhuh
[with so] a crest that we on huh[hang] huh
[with li]ke a crest that we canhuh[hang] huh
[ ]
030 E [huh:]
[huh:]
[huh:]

031 M köhuhnnen,
cahuhn,
onhuhit,

032 (0.9)

033 M °huh huh°
°huh huh°
°huh huh°

034 E °·huh°=
°·huh°=
°·huh°=

035 M =·h:: oh je oh je oh je¿
=·h:: oi ye oi ye oi ye¿
=·h:: oi ye oi ye oi ye¿

036 E h:: huh huhuh,
h:: huh huhuh,
h:: huh huhuh,

037 (0.4)

038 E ·h::=
·h::=
·h::=

039 M =>↑ich weiß nicht< ob die brunhild das so gut
=> ↑i know not < if the brunhild that so good
=> ↑i don't know< if brunhild would think much

040 findet.
finds.
of that.

041 (0.1)

042 M huh:=
huh:=
huh:=

043 E =huh:
=huh:
=huh:

044 (0.1)

045 E ach so so gegen so en ↑stAmmtisch↓aschenbecher,
ach so so against so a ↑rEgulars' table↓ash tray,
ach so well surely she has nothing against like,

046 M °·h: [huhuh° stimmt]
°·h: [huhuh° is right]
°·h: [huhuh° true]
[ ]
047 E [hat sie bestimmt nichts]
[has she surely nothing]
[a ↑stAmmtisch ↓ash tray]

048 M die andere is- ist doch auch ↑bei jAnas vAter
the other is- is still also ↑by jAna's fAther
the other is- is also still over ↑at

049 beim stAmmtisch mit dabei¿
by the regulars' table with thereby¿
jAna's fAther' stAmmtisch there¿

050 (0.2)

051 E ach gott. stimmt,
oh god. is right,
oh god. that's right,

052 (0.6)

053 M °h::° HUH huhuh
°h::° HUH huhuh
°h::° HUH huhuh

054 (0.9)

055 M ·huh·h:: oh je. oh je,
·huh·h:: oi ye. oi ye,
·huh·h:: oi ye. oi ye,

056 (2.2)

((Begin Excerpt 7))

057 M die brunhild ↑mA:lt ja ↑auch¿
the brunhild ↑pAI:nts yes ↑also¿
brunhild ↑pAI:nts too ↑right¿

058 (1.3)

059 E Ach stImmt die hat uns ja auch schonmal so
Oh is rIght the has us yes also already once so
Oh that's rIght she already put like

060 so so en pamphlEt von irgendso einer
so so a pamphlEt of some sort of an
like like a pAmphlet about some sort an

061 AUsstellung <° aufn tisch gelegt.°>
Exhibit <°on the table placed.°>
Exhibit <° on our table. °>

062 M °·h:: ptch° ja?
°·h:: ptch° yea?
°·h:: ptch° yea?

063 (0.5)

064 M hath mir auch irgendwie was von ner AUsstellung
hash me also somehow what of an Exhibit
alsoh told me something about an Exhibit

065 er↑zählt wo sie jetzt bAld ihre bIlder ausstellt,
t↑old where she now sOOn her pAIntings shows,
where she is ↑showing her pAIntings now sOOn,

066 (1.2)

067 E da:: ↑kommste wohl nicht ↑drum ↑rum.
there:: ↑come you probably not ↑around ↑round.
that:: ↑you're probably not getting ↑out ↑of.

→ **068 M** ↑nee das ist gAnz wEIt wEg;= da komm ich ☺
↑nah that is qUite fAr awAy;=there come i ☺
↑nah it is sO fAr awAy; =i can't get over ☺

→ **069** nicht hin. ☺
not there. ☺
there. ☺

((End Excerpt 7))

070 (0.2)

071 E huh[uhuhuhuhuhuhuh huhuh ·h::]
huh[uhuhuhuhuhuhuh huhuh ·h::]
huh[uhuhuhuhuhuhuh huhuh ·h::]
[ ]
072 M [huhuhuh h:: ↑tut mir ↑leid,] ☺
[huhuhuh h:: ↑does me ↑sorrow,] ☺
[huhuhuh h:: ↑i am ↑sorry,] ☺

```
073        (0.3)

074  E     °>·h:<° ↑Oh::   wie  ↑schA↓de, huhuh ☺
           °>·h:<° ↑Oh::   how  ↑shA↓me,  huhuh ☺
           °>·h:<° ↑Oh:: what a ↑shA↓me,  huhuh ☺

075  M      hja,  aber¿ huh
           hyeah, but¿  huh
           hyeah, but¿  huh

076        (1.3)

077  M     ·h::
           ·h::
           ·h::

078  E     °huhuh[uhuh°]
           °huhuh[uhuh°]
           °huhuh[uhuh°]
                 [     ]
079  M          [n    ]ein;
                [n    ]ein;
                [n    ]ein;

080        (0.8)

081  E     huhuhu[huh          ]
           huhuhu[huh          ]
           huhuhu[huh          ]
                 [             ]
082  M          [>ich  glaub <] den philipp hat sie aber
                [> i  believe<] the philipp has she but
                [>but i think<] she Actually Always left

083        EIgentlich Immer  zIEmlich in ruhe  gelassen,
            Actually  Always  rAther  in quiet   left,
                    philipp prEtty mUch alone,
```

As Marion completes her telling of the work to be done the next day in her new apartment, Elena produces an assessment of these events as being "exciting" ("*aufregend*"; line 2). This adjective is used as a sarcastically negative assessment to describe the sort of excitement and general dishevelment one might associate with a day of painting a room, and indeed it is the last in a series of negative assessments of the busy day to come, in which the

painting of decorative stripes is portrayed as marginally less awful than the work that one normally associates with painting a room. In line 4, Elena then proffers a new topic by assessing what the new apartment will be like once Marion moves in, stating that they will be living in a "dive bar" ("*Spelunke*"; 4). Marion repeats the latter half of this utterance in a quiet, breathy voice in line 5, while Elena begins to laugh in overlap (6), treating her own previous assessment as non-serious and inviting co-participant laughter from Marion, an invitation which is then accepted (8).

After a gap of 0.4 seconds, Marion produces a quiet partial repeat of Elena's utterance before stating that there could be a "*Spelunkenrat*" (10). This term, as such, does not exist in any standard German dictionary, but is a compound noun derived from the German words "*Spelunke*" (dive bar) and "*Rat*" (advice, or advisory council), as is common in German. Due to the rather idiosyncratic nature of the term, I took the unusual step of following up with Marion to inquire about its meaning, without providing the context in which it arose. She recalled the term in isolation, as she had taken up regular use of it subsequent to this recording. Marion stated that she was describing how her friends often congregate at her home in the kitchen and discuss life, and was therefore suggesting in a tongue-in-cheek manner that she and her new roommate could formalize this practice into a sort of kitchen dive bar advice or life coaching service.

In response to this suggestion that they formalize Marion's new home into a *Spelunkenrat*, Elena indicates that she and her friends will give Marion a "*Stammtischaschenbecher*" (19) to place on her table (21 and 23). This is a special type of ash tray (*Aschenbecher*), usually with a small sign above it, which is traditionally used to mark a table as reserved for a regularly occurring group meeting (*Stammtisch*), typically associated with older men. This would presumably be intended as a housewarming gift which would encourage

the use of Marion's table as the new *Spelunkenrat*. As Elena describes the gift, she treats her suggestion as intrinsically humorous by producing laugh tokens embedded in her speech, however, the suggestion is treated as serious by Marion, who assesses it as "amazing" ("*toll*"; 22). In other words, while the idea of a special ash tray to identify the newly founded *Spelunkenrat* is treated as intrinsically humorous by the participants (i.e., it is treated as a laughable), the serious import of suggesting such a gift is receipted as such by means of positive assessment. Indeed, when queried about the meaning of *Spelunkenrat*, Marion also stated that such an ash tray was in fact gifted to her after she moved into her new apartment, supporting the serious interpretation of this spate of talk, despite the infusion of laugh tokens. Furthermore, given that both Elena and Marion smoke cigarettes, such an ash tray would presumably be a useful item.

In lines 24 and 27, Marion expresses a further wish that the ash tray have a sign on it, where they could then hang a crest or coat of arms ("*Wappen*"; 29). Marion and Elena again both treat this wish as intrinsically humorous. Marion produces embedded laugh tokens in the latter half of her utterance (29 and 31), while Elena laughs in overlap and subsequent to Marion's utterances (30, 34, and 36). However, they also treat the action of suggesting the procurement of such an item as serious. In lines 39-40, Marion indicates that she is unsure if her new roommate, Brunhild, will like the proposed ash tray with sign, to which Elena responds by stating her belief that Brunhild should have no problem with such an item (45 and 47). These utterances reveal that, while the idea of such an ash tray is treated as humorous by them, they are also considering the serious implications of obtaining one and placing it in a shared space. Therefore, we cannot consider the utterances to be non-serious in the way in which Sacks (1972) defined non-seriousness as making irrelevant any subsequent action besides laughter.

In line 48, Marion indicates that "the other one" ("*die andere*"; 48) participates in the *Stammtisch* of a friend's father. Unfortunately, it is not clear from the context what this references, although Elena confirms this information is "right" ("*stimmt*"; 51). At this point, Marion produces a series of laugh tokens (53) and the tokens "*oh je oh je*" (55), which bring the preceding sequence to a close.

After a gap of 2.2 seconds, Marion then proffers a new topic in the form of a telling, namely that her new roommate Brunhild paints (57). Elena receipts this information with the token "*ach*" and the lexical item "right" ("*stimmt*"; 59), indicating that she had previous knowledge of this (a fact which was acknowledged by Marion in her utterance with the token "*ja*"; 57). She then provides grounds for such previous knowledge by telling Marion that Brunhild had already given her a pamphlet about an upcoming exhibit (59-61). Marion utters the acknowledgment token "*ja*" in line 62, and then, after a gap of 0.5 seconds, indicates that Brunhild had also told her about an exhibit where she would be showing her work soon (64-65).

Following a gap of 1.2 seconds, Elena indicates that the exhibit is something Marion is "probably not getting out of" ("*da kommste wohl nicht drum rum*"; 67), and has no choice as to whether she will attend or not, regardless of her level of interest in her roommate's artwork. Marion refutes this by stating that the exhibit is really far away and that she cannot go there (68-69). This refutation of Elena's assertion is uttered with no gap between the two utterances. Given that such a refusal is normally a dispreferred response (Pomerantz, 1984), which would then be accompanied by such early indications of a forthcoming dispreferred action, such as gaps, hesitations, or other discontinuities in the talk (Sacks, Schegloff, & Jefferson, 1974), this is an early indication that perhaps this utterance is doing some other action.

Elena treats this refusal as non-serious by laughing (71). Marion joins in her laughter

(72), treating her own utterance as non-serious. She then apologizes with, "I am sorry" ("*tut mir leid*"; 72) in a sing-song voice, which, in conjunction with the laughter and the previous non-serious utterance that the exhibit is too far away, can also be interpreted as non-serious. In other words, Marion is here performing hypothetical talk (Golato, 2012) for Elena about what she would say to Brunhild in order to avoid attending the exhibit, while simultaneously treating her excuse and expression of regret as non-serious in the physical absence of Brunhild.

To this point, both Marion and Elena have treated Marion's excuse as non-serious by producing only laughter and pursuing no other potentially relevant second action (Sacks, 1972). At line 74, after a gap of 0.3 seconds, Elena produces the utterance "what a shame" ("*wie schade*"). This may be interpreted as a non-serious hypothetical performance of Brunhild's potential reaction to Marion's excuse (which builds on Marion's performance; Golato, 2012), or as a serious negative assessment of the situation itself (i.e., that Marion is not able to attend Brunhild's exhibit), however, Elena herself treats this utterance as non-serious, or ironic, by laughing (74).

Marion does not initially laugh in response. Rather, she makes the utterance "yeah but" ("*ja aber*"; 75), followed by laughter and "no" ("*nein*"; 79). The return to seriousness is then accompanied by only a very slight shift in topic, similar to what we observed in Excerpt 5/6. Marion indicates that Brunhild left Philipp, Brunhild's previous roommate and a friend of Marion and Elena's, to his own devices (82-83). Talk then continues with Marion expressing her hope that Brunhild will do the same to her, while also expressing doubts that this will be case.

The above analysis demonstrates that lines 68-69 are treated as non-serious by the participants in this conversation, which stands in contrast to their serious treatment of Marion's utterance in lines 24 and 27. These utterances are therefore included in the collection for

acoustic analysis and perceptual testing.

3.2.2. Results: Sequential Analyses

The above analyses demonstrate that these excerpts all contain instances of non-seriousness according to the CA definition outlined above. Specifically, the utterances are treated as non-serious by both conversationalists in that they are followed by laughter, with no serious uptake of the utterances in the form of a relevant second action besides the laughter. Furthermore, talk continues after the laughter with no signs of initiation of repair, i.e., neither the laughter nor lack of serious response are treated in any way by either participant as a source of trouble in need of repair. Finally, subsequent to each utterance, there is a jointly managed return-to-seriousness. We also observed that these instances of non-seriousness are all preceded by spates of talk which are treated as serious by the participants in the conversation. These utterances could therefore be harvested for use as controls in the acoustic analyses presented below.

Finally, all of the utterances included in this final collection fit the criteria for inclusion in the perceptual survey, as detailed in Chapter 2. Specifically, they can be shown through sequential analysis to be treated as non-serious, they are uttered in a voice which differs perceptibly from the surrounding talk, and they are not uttered in overlap with any laugh tokens or other utterances from the co-participant in the conversation, such that it was possible to prevent participants in the perceptual survey from knowing whether or not they laughed in response. The results of this perceptual experiment are presented in Chapter 4.

3.3. Acoustic Analysis and Non-Seriousness

Having selected the excerpts for acoustic analysis on the basis of the sequential analyses presented above, each utterance of interest was trimmed from the .wav file. While the CA work

was conducted utilizing the .mp4 format video data, the acoustic analyses were done using the .wav format audio data which was collected concurrently with, but separately from, the video data (see Chapter 2). The use of separate, concurrently collected data streams allowed for superior analyses, both with CA and with PRAAT. If technology permits, CA methodology encourages the use of video data when interlocutors are in each other's physical presence, while accepting audio with a lower sampling rate and as a result, with a reduced pitch range. Conversely, acoustic analyses are best conducted on audio samples which contain the full pitch range produced by humans. The .wav file format collects this full range of pitch at a higher sampling rate than is possible to capture with the .mp4 video or .mp3 audio formats. Therefore, the acoustic analyses described below were conducted on excerpts trimmed from the original .wav files in order to capture the fullest range of linguistic detail possible when using digital media.

As noted above, a total of eight excerpts from four unique speakers were subjected to further acoustic analysis. For each speaker, an utterance treated as non-serious was located, in addition to an utterance treated as serious from talk immediately preceding the non-serious utterance. The utterances were trimmed at the points indicated by arrows (→) in the transcripts above, such that the full utterance or intonational unit was considered intact. Two of the speakers were male, and two female. All speakers spoke relatively standard varieties of colloquial High German, inflected with some regional variations in pronunciation and word choice. Two of the speakers spoke with regional hallmarks of Hessian, and two with hallmarks of Berlin-Brandenburgish (see Table 3.1., below, for a summary of the demographics for each excerpt). All speakers spoke German as their native language, were Caucasian, and were enrolled in a university degree program at the time of recording.

Table 3.1.

Speaker Demographics		
Speaker	**Gender**	**Regional Variety**
Simone (E 1/2)	female	Berlin-Brandenburgish
Maximilian (E 3/4)	male	Hessian
Gerhart (E 5/6)	male	Berlin-Brandenburgish
Marion (E 7/8)	female	Hessian

As Repp (1981a, 1981b, 1983a, 1983b, 1983c) argues, a given phonological contrast is cued by multiple articulatory features which operate under a set of trading relations. This multiplicity of articulatory cues allows that the phonological contrast, while normally cued by a *set* of articulatory feature manipulations, may in fact still be perceptually achieved in the absence of one or more of its normative cues. For example, contextual affects such as coarticulation may disrupt or inhibit perception of one of the normally present articulatory cues giving rise to a given phonological contrast, yet contrast is preserved by production and perception of the other articulatory cue(s).

In consideration of Repp's (1981a, 1981b, 1983a, 1983b, 1983c) theory of phonetic trading relations, it was presumed from the outset that any acoustic marking of an utterance as non-serious would likely consist of manipulation of a constellation of acoustic features, as is implicit in Hypothesis 1A, above. Given the status of this study as exploratory in nature, if multiple non-serious utterances in the set show evidence of manipulation of a given feature, this would indicate that further study of that particular feature with a larger data set in order to determine statistical significance would be warranted, while others may be discarded as showing little need for further study. However, it may also be the case that not all of the manipulated acoustic features must be copresent for an utterance to be successfully perceived and treated as

non-serious, as these features may represent a constellation of cues in a phonetic trading relation.

In other words, a given acoustic feature need not necessarily be manipulated by all speakers in the data set in order to be considered a potential candidate for further study. Rather, the results of the acoustic analysis presented here must also be considered in light of the results of the perceptual experiment presented in Chapter 4. The acoustic features of the non-serious utterances which were also most consistently identified as such by participants in the perceptual experiment are given particular weight in the final integrated analysis presented in Chapter 5, and therefore are also those features which should be considered in any future acoustic analyses of larger data sets.

The selection of the acoustic features measured for this study was based on an extensive review of the literature available to date on the acoustics of humor. The majority of these studies were conducted on data collected in a laboratory or from scripted sources, such as television and radio programs. The results of the present study are therefore not directly comparable to the results of previous research due to the major differences in methodologies, however, previous research laid the important groundwork on which this study builds by providing a basic set of potential phonetic cues to consider. This set of articulatory features consists of: pitch, pitch variation, vowel formants, intensity, beats, and speech rate. The results of previous studies for each feature are reviewed in the relevant sections throughout the rest of the chapter. Due to the small size of the data set, no claims about statistical significance are made, rather the conclusions are based on observation of general trends in the data which are suggestive of further investigation with a larger data set.

3.3.1. Pitch

The first set of analyses conducted on the excerpts examined pitch, or fundamental

frequency (F0). This required a two stage process, wherein pitch tracks were first created and then compared to the automatically generated pitch measurements produced by PRAAT. This two-step process was necessary because it allowed for machine errors to be readily identified and eliminated from consideration, as well as resulting in a series of pitch tracks suitable for visual analysis of variation in pitch, as described below.

For each excerpt, a pitch track was created in PRAAT, with the pitch floor set at 60 Hz and the ceiling at 350 Hz for all speakers (Vogel, Maruff, Snyder, & Mundt, 2009). The resultant track was then smoothed at a bandwidth of 10 Hz. This smoothed pitch track was then visually inspected with reference to the spectrogram of the original data excerpt. Local pitch peaks and valleys were identified and marked in a text grid aligned with the spectrogram. Additionally, as a result of this visual inspection, the likely location of the maximum and minimum pitch for each excerpt was identified by hand and annotated in the spectrogram.

Using the same pitch floor and ceiling of 60 Hz and 350 Hz, respectively, and with auto-correlation enabled, the maximum, minimum, and average pitches were measured in PRAAT. The maximum and minimum pitches were first located automatically with the program, and then visually and acoustically inspected, in conjunction with the annotated text grids, to detect any machine errors. In this way, it was possible to eliminate from consideration false spikes in pitch associated with fricatives, and also with the chirping of birds in one instance (Excerpt 2), in order to locate and measure the true maximum pitch. The average pitch was measured across the entirety of the trimmed excerpts, excluding any surrounding silences or ambient noise.

Previous studies which measured manipulation of pitch as a potential acoustic feature of humor or non-seriousness have had conflicting results. Of the studies that reported a clear trend in manipulation of pitch, most claimed to find that speakers consistently increased the pitch of

humorous utterances, including three of the five studies located which made use of naturally occurring conversational data (Ford & Fox (2010), Hidalgo Navarro (2011; Spanish[2]), Kohler (2008; German). However, Hidalgo Navarro's (2011; Spanish) study does not provide any measurements of pitch values to support the analysis, while Ford and Fox' (2010) study does so only sporadically. On the basis of laboratory data using [hVd] syllables, Tartter (1980) and Tartter and Braun (1994) also report an increase in pitch of humorous utterances.

In contrast, Bird (2011), Pickering, Corduas, Eisterhold, Seifried, Eggleston, and Attardo (2009), Rockwell (2000), and Voyer and Techentin (2010) all report a decrease in pitch. Bird's (2011) study made use of naturally occurring data, and therefore stands in direct contrast to the studies described above. The latter three utilized script data, were participants were asked to perform a canned joke. Finally, Attardo, Pickering, and Baker (2011), Auberge and Cathiard (2003; French), and Flamson, Bryant, and Barret (2011) reported no change in pitch from baseline. The results from Flamson et al. (2011) were based on naturally occurring data, while Attardo et al. (2011) collected conversational data in a laboratory setting. Auberge and Cathiard (2003; French) collected sentences read from a script.

Other studies have found that manipulation of pitch in humorous or non-serious utterances varies depending on additional variables. Anolli et al. (2000; Italian) differentiate between sarcastic and kind irony using script data, the former marked by an increase in pitch while the latter exhibits less change from baseline values. Attardo et al. (2003) found both increased and decreased pitch in humorous lines collected from television situation comedies. Bryant and Fox Tree (2005) report an increase in pitch from baseline values for unambiguously ironic utterances collected from radio data, while ambiguously ironic utterances did not vary

[2] Except where otherwise noted, all studies used English language data.

significantly from baseline. Using conversations collected in a laboratory, Bryant (2010, 2011) found that manipulation of pitch varies by speaker, with some increasing and others decreasing the pitch of humorous utterances. In a series of studies, Cheang (2008) and Cheang and Pell (2008, 2009, 2011) document an increase in pitch in humorous as well as sarcastic utterances in Cantonese, while showing that English speakers decrease their pitch for comparable utterances. Finally, Purandare and Litman (2006) and Urios-Aparisi and Wagner (2011) both find increases and decreases in pitch of humorous utterances collected from the television series *Friends* and *Sex and the City*, respectively, with results varying by gender of speaker and context.

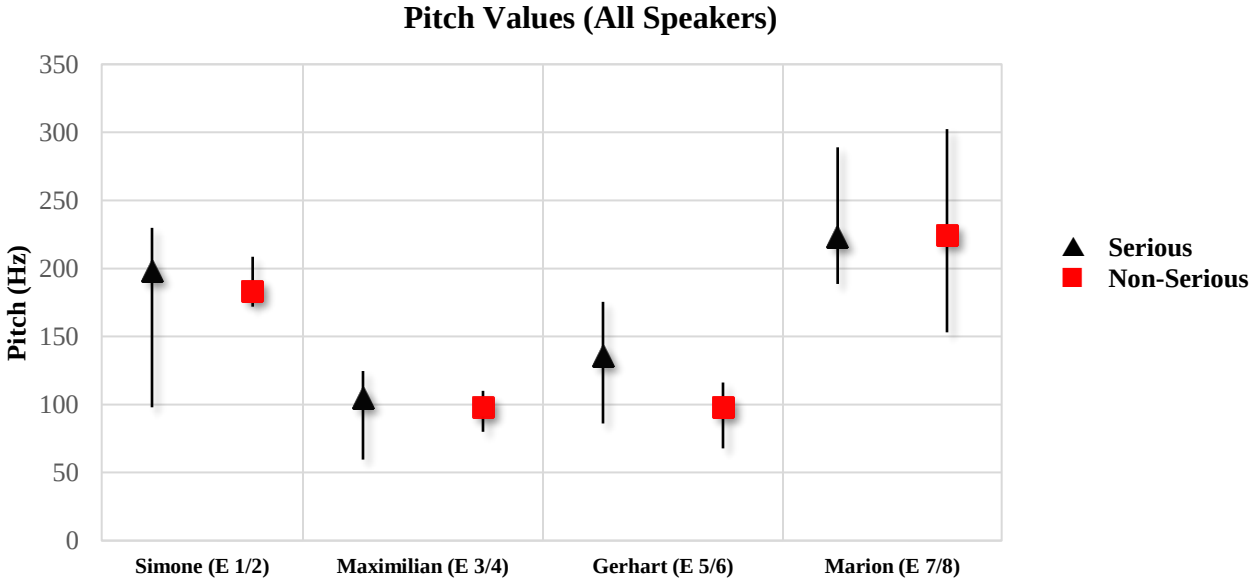
On the basis of the widely divergent results from the literature cited above and the pilot version of this study, it was anticipated that there would not be a uniform trend towards raising or lowering of pitch. In the pilot study, which used English data, participants had not been found to reliably increase or decrease their maximum, minimum, and/or average pitch in order to mark non-serious utterances as such. Rather, they consistently modulated their pitch in a manner which deviated from that of the surrounding talk. In particular, this modulation of pitch resulted in a narrowing of the pitch range, as well as in a convergence of pitch ranges across speakers. In other words, when making identifiably non-serious utterances, speakers had a smaller range of pitch which was more similar to that of other speakers making non-serious utterances than to samples of their own surrounding serious speech. Such an interpretation would fit neatly with a more concise view of the previous work in the area, since whether speakers show an increase, a decrease, or no change in pitch would naturally vary according to how their baseline pitch compared to the pitch range of non-seriousness. It was anticipated that speakers of German might also similarly have a reduced pitch range which converged with that of other speakers when making non-serious utterances. As argued below, the data reveal that this anticipated

result does indeed hold for naturally occurring German conversational data.

3.3.1.1. Data: Pitch

Figure 3.2., below, contains a summary of the pitch maximum, minimum, and average values collected for all speakers.

Figure 3.2.

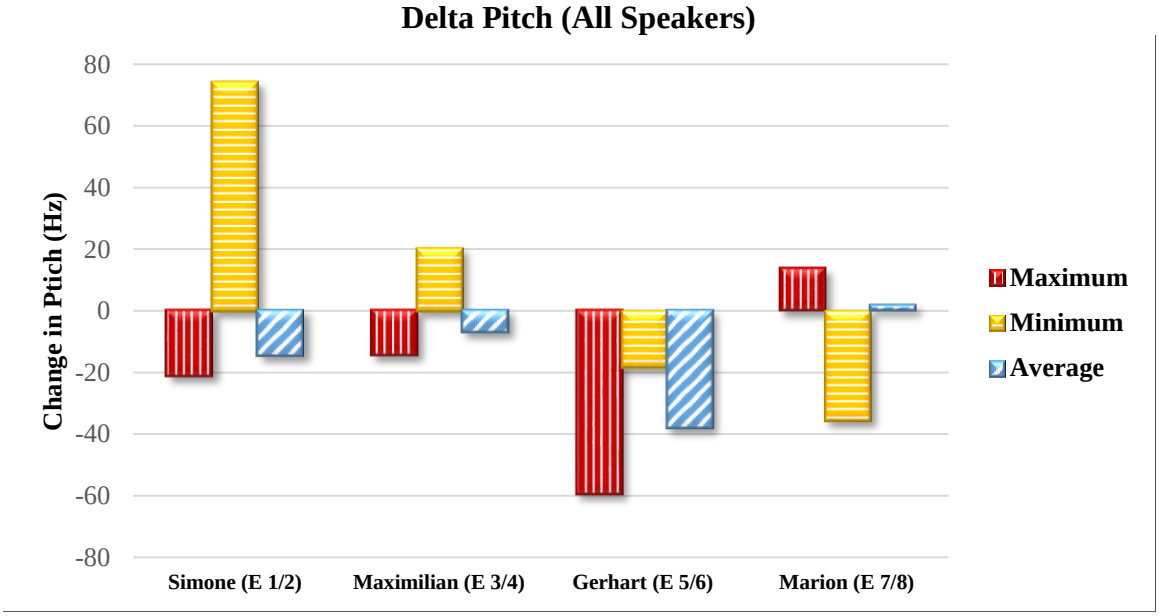


Speaker	Maximum		Minimum		Average	
	▲	■	▲	■	▲	■
Simone (E 1/2)	229.82	208.58	97.99	172.02	198.26	183.56
Maximilian (E 3/4)	124.62	110.18	59.55	79.87	105.12	98.04
Gerhart (E 5/6)	175.46	116.14	85.99	67.79	135.58	97.67
Marion (E 7/8)	288.93	302.39	188.61	153.06	222.93	224.58

As can be seen in the data above, there was no consistent pattern of increasing and/or decreasing of pitch values in order to mark non-seriousness. When making non-serious utterances, Simone, Maximilian, and Gerhart had lower maximum pitch values in comparison to adjacent serious utterances, while Marion had a higher maximum pitch. The minimum pitch values of non-serious utterances were lower for Gerhart and Marion, but higher for Simone and Maximilian.

Finally, the average pitch of non-serious utterances was lower for all speakers, except Marion, who had a slightly higher average pitch. Figure 3.3. summarizes the delta values of the raw pitch measures for all speakers, where a positive value indicates an increase in pitch from serious to non-serious talk and a negative value a decrease in pitch.

Figure 3.3.

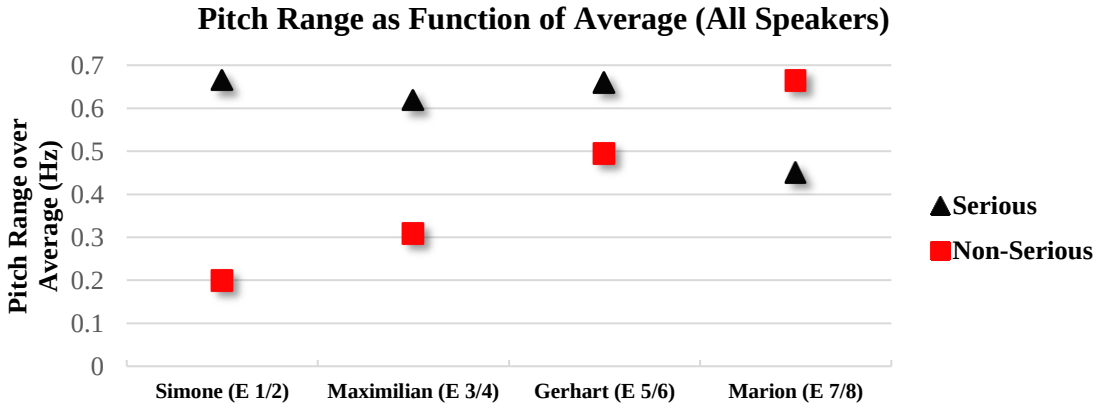


While there is no across-the-board trend in terms of lowering and/or raising of pitch, it should be noted that there is a trend towards a reduction of the range of pitch when shifting from serious to non-serious talk. Returning to Figure 3.2., we see that three of four speakers had a reduced pitch range when making non-serious utterances in comparison to their serious utterances, a result similar to what was found in the pilot study, as noted above. In addition, we note that the pitch ranges for the male speakers, Maximilian and Gerhart, became more similar during non-serious talk. This convergence of pitch patterns does not hold for the female speakers, due to Marion having an increased pitch range for her non-serious utterance, however, as described below, Marion's speech contains other potential markers of non-seriousness. Unfortunately, a comparison of the male and female speakers in terms of pitch range

convergence is fruitless, due to the physiologically determined differences in pitch range between male and female speakers. Because male speakers tend to have a lower pitch range than female speakers anyway (Vogel et al., 2009), such convergence of pitch ranges across the sexes would be unlikely to occur unless a given male speaker had an unusually high pitch range and/or a given female speaker had an unusually low pitch range, which is not the case here. We therefore can only observe pitch range convergence in this data set across the two male speakers.

When viewed as a function over the average, the reduced pitch ranges of the speakers reveal a tendency for reduced variation of pitch over the course of the utterance. In other words, speakers are exhibiting a trend towards tighter control over the pitch of their speech when making non-serious utterances, while relaxing this control over the course of spates of serious utterances. As the speakers place tighter control over the pitch range, the result is both the narrowing and converging of pitch ranges, especially as seen with the male speakers in these data. Figure 3.4. contains a visual representation of the trend towards a reduction in variability of pitch over the course of non-serious utterances as compared to serious utterances.

Figure 3.4.



Here, lower values correspond to reduced variability in pitch. As can be seen, three of the four speakers exhibit evidence of reduced variability in pitch when making non-serious utterances in comparison to serious utterances.

3.3.1.2. Results: Pitch

As presented above, the data reveal no clear observable pattern of increasing and/or decreasing of the maximum, minimum, or average pitch values in order to mark non-seriousness. However, speakers did modulate these pitch values in such a manner as to have a reduced pitch range. Additionally, there was some evidence that the pitch range converged towards a "non-serious" pitch range which was similar across speakers, especially for the male participants, such that their pitch ranges became more similar when making non-serious utterances as compared to serious ones. Finally, we observed a trend for reduction of the variation in pitch over the course of the utterance during non-serious talk.

Excerpt 7/8 can be viewed as an outlier, in the sense that Marion, at least in terms of pitch, behaves differently from the other speakers. That is to say, she does not take advantage of the option to modulate her pitch range to mark non-seriousness, whereas the other speakers do. While Marion does not utilize this strategy in order to mark her utterance in Excerpt 7 as non-serious in contrast to Excerpt 8, as we will see below, she does mark the utterance using other features available to her in the linguistic toolkit.

On the basis of this analysis of the data, it can be concluded that modulation of the pitch range, and especially a narrowing of the pitch range towards an as-yet undefined target range, is likely a marker of non-serious in contrast to serious utterances. Pitch can therefore be considered a member of the constellation of features which may be manipulated in order to mark turns at talk as non-serious (i.e., as laughables) and which is need of further study.

3.3.2. Pitch Variation

In the process of measuring speakers' pitch, a series of pitch tracks were created in order to visually inspect the movement of their pitch over the course of the excerpts. As described

above, the tracks were created using a pitch floor and ceiling of 60 and 350 Hz, and then smoothed at a bandwidth of 10 Hz. The resulting tracks were then drawn in overlap using PRAAT, which enabled a visual comparison of the changes in pitch between serious and non-serious utterances.

In English, unemphatic, declarative sentences are generally uttered with an initial peak in pitch towards the beginning of the utterance and a final drop in pitch (Ladefoged, 2001). There is a general trend, in other words, to move from higher to lower pitches, albeit with pitch peaks at intervals within the intonational unit. German has a similar pattern of rise-fall intonation in declarative sentences with local internal pitch peaks and valleys (Hirst & Di Cristo, 1998).

Of the studies of the acoustics of humor, few specifically considered variation in pitch. Using naturally occurring data, Bird (2011) found a decrease in variation in pitch of non-serious utterance, while Flamson et al. (2011) documented no significant change. On the other hand, Bryant (2010) concluded that there was an increase in variation in pitch using conversational data collected in a laboratory setting. Results from reading script data were also mixed, with Anolli et al. (2000; Italian) finding an increase in variation and Cheang and Pell (2008, 2009, 20011; Cantonese and English) and Rockwell (2000) finding a decrease. Finally, using radio data, Bryant and Fox Tree (2005) concluded that there was no change in variation of pitch, while Purandare and Litman (2006), using television data, found an increase in variation. Also using television data, Attardo et al. (2003) found both increases and decreases in variation, depending on the content of the utterance.

It was anticipated prior to the experiment that the general pitch pattern of non-serious utterances would differ from that of serious utterances. In particular, I hypothesized that there might potentially be a final local rise in intonation instead of the normal fall. Although this

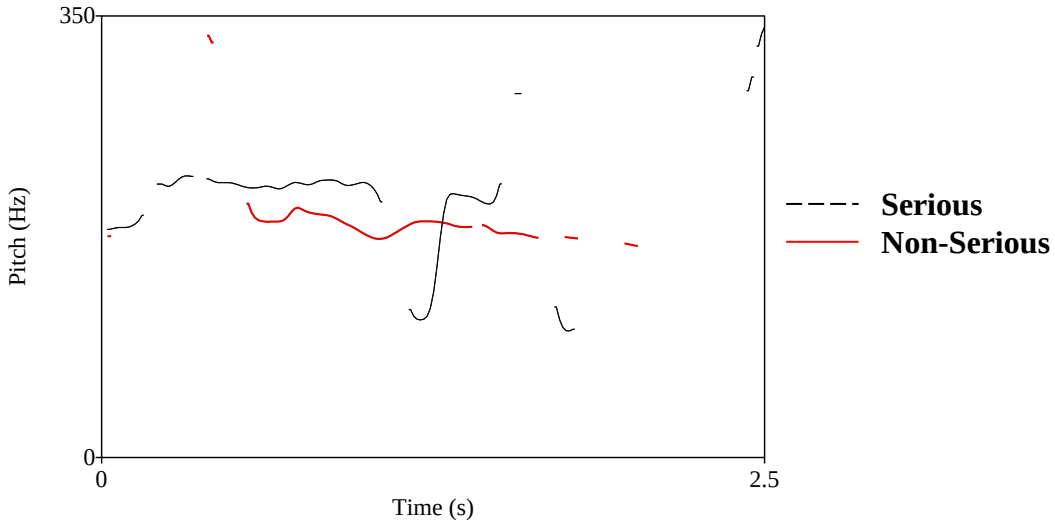
particular hypothesis was not borne out by the data, the results of a visual inspection of the pitch tracks fit well with the pitch data described above, and that non-serious utterances do, in fact, follow a slightly different pitch pattern than serious ones.

3.3.2.1. Data: Pitch Variation

The pitch tracks for each speaker are presented below in a series of overlapping images, such that the serious and non-serious utterances are mapped into the same space to allow a visual comparison of the two. Figure 3.5., below, contains the pitch tracks for Simone (Excerpts 1/2).

Figure 3.5.

Pitch Track: Simone (Excerpts 1/2)



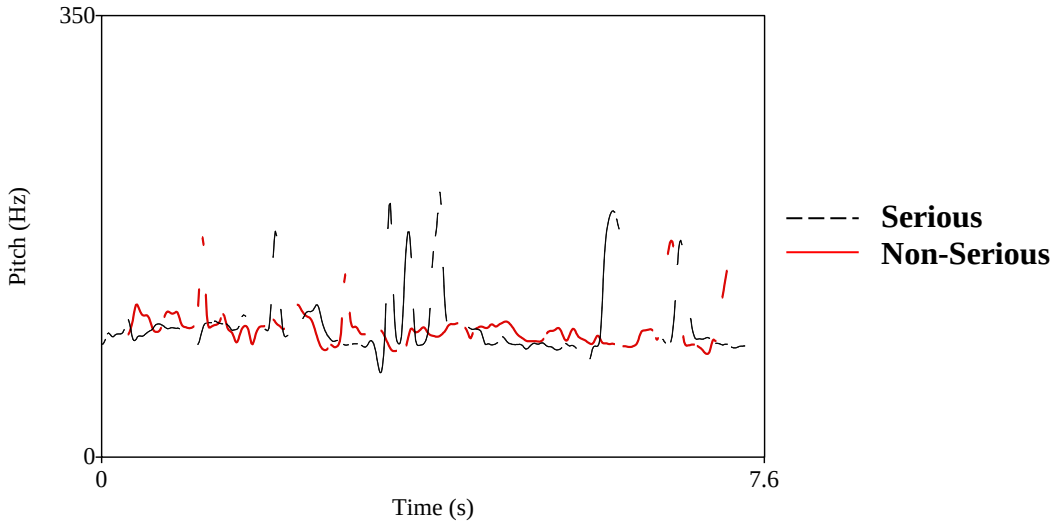
Here, we can see that Simone has a rise-fall intonation pattern for both her serious and non-serious utterances (the extreme pitch excursion at the end of the serious utterance is attributable to the chirp of a bird through an open window). While there are pitch peaks and valleys throughout the utterances, the general trend is a falling one over the course of both utterances. However, we note also that the peaks and valleys in the non-serious utterance are less extreme in comparison to those in the serious utterance, in that the peaks are lower and the valleys higher in the serious excerpt. This fits well with the results from the previous section, where we observed

that there was a tendency for reduced variability in pitch in non-serious utterances. This then visually confirms those earlier results.

The pitch tracks for Maximilian (Excerpts 3/4) are presented below in Figure 3.6.

Figure 3.6.

Pitch Track: Maximilian (Excerpts 3/4)

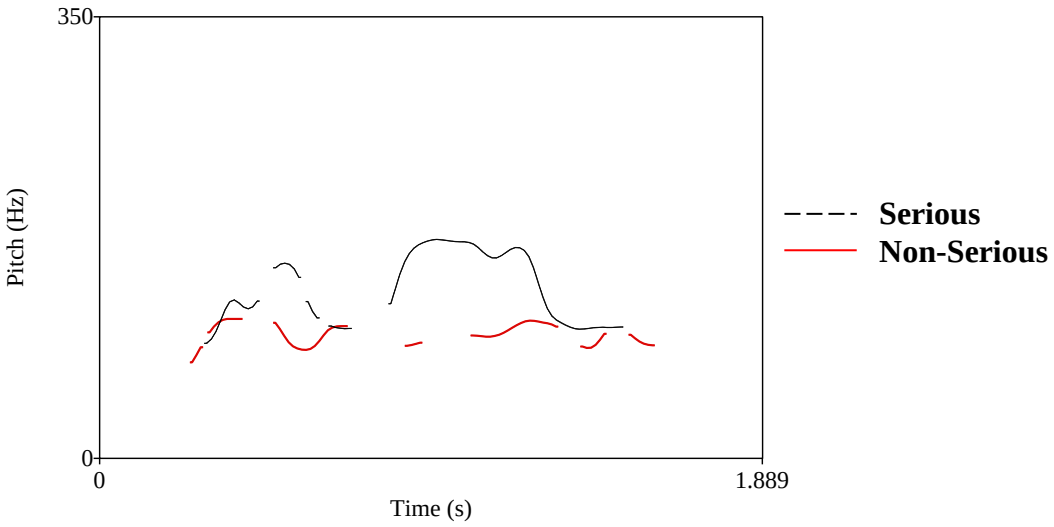


Similar to Simone, Maximilian's data reveal a trend towards falling intonation over the course of both utterances. While there appears to be a final peak in pitch at the end of the non-serious utterance, however, this is in fact associated with a laugh token produced by Emmanuel, Maximilian's co-conversationalist, and therefore should not be considered a part of Maximilian's pitch track. Again as in the previous example, we note here that the pitch peaks and valleys are slightly less extreme when making the non-serious utterance in comparison to the serious utterance, especially in the latter half of the utterances.

Figure 3.7. contains the pitch tracks for Gerhart (Excerpts 5/6).

Figure 3.7.

Pitch Track: Gerhart (Excerpts 5/6)

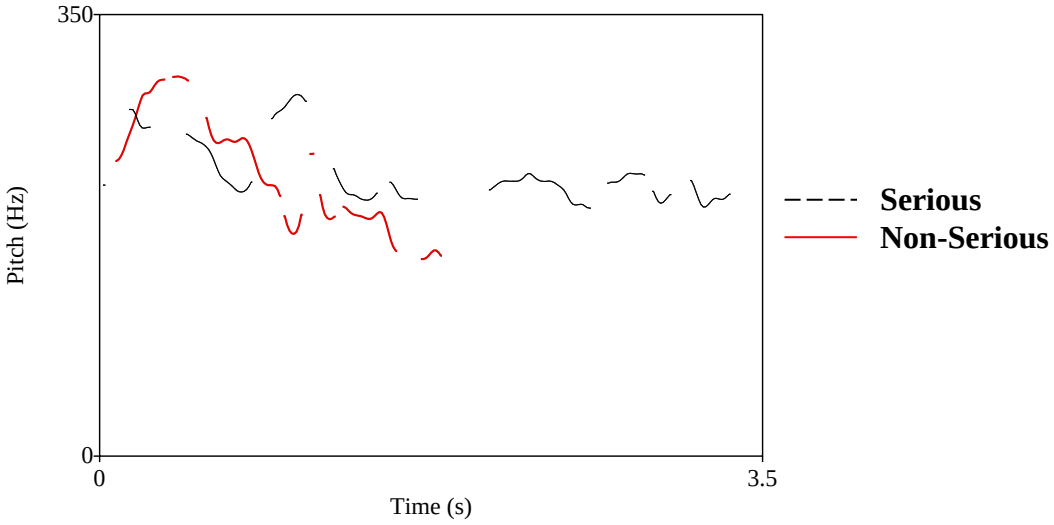


As in the previous excerpts, Gerhart's pitch tracks reveal a rise-fall pitch pattern in both non-serious and serious utterances, with a general tendency for a falling pitch over the course of the entire utterance. Note also that the non-serious utterance again has less extreme peaks and valleys than the serious one, indicating reduced variation in pitch over the course of the former.

Finally, we turn to Marion's data below in Figure 3.8.

Figure 3.8.

Pitch Track: Marion (Excerpts 7/8)



Marion's pitch tracks are similar to those viewed above in that they also reveal a similar rise-fall pitch pattern with an overall falling trend. However, they differ in one important aspect, namely, the pitch range is larger, as we would expect based on her pitch data presented in the previous section. The movement from the initial rise to the next local fall in the non-serious utterance is greater than in the serious utterance, although the pattern of peaks and valleys in the latter half of both utterances is very similar, with similar degrees of variation. In this way, Marion's speech differs from the others' in terms of changes in pitch of non-serious utterances, a finding which is again in line with the results from the previous section.

3.3.2.2. Results: Pitch Variation

The results of a visual analysis of the pitch tracks created for this set of speakers reveal that the pitch patterns of non-seriousness differ from those found in serious spates of talk. As stated above, it had been hypothesized that this would be the case, and that there might potentially be a final local rise in pitch associated with non-seriousness. However, this hypothesis was not borne out by the data. In the sole non-serious excerpt (3) which contained a final local rise in pitch, this was associated with laughter from the co-conversationalist in response to Maximilian's utterance.

Despite this, there was an observable difference in the pitch patterns of non-seriousness as compared to seriousness. Specifically, the rise-fall pitch pattern and general falling pitch with internal peaks and valleys prototypical of unemphatic declarative sentences was found to hold for non-serious utterances as well, however they also exhibited reduced variation in pitch over the course of the utterance. This can be seen in that, in the speech of three of the four participants, while there is still local variation in pitch, the local peaks are relatively lower and local valleys are relatively higher in pitch in the non-serious utterances than in the serious ones.

Marion's (Excerpts 7/8) speech again differs from the others', as she does not modulate her pitch pattern in any observable manner in order to differentiate between seriousness and non-seriousness. This result is not surprising, considering that her pitch values presented above also differed in a similar way from the other excerpts, i.e., in her non-serious utterance she did not have a pitch profile which was markedly different from her serious one. Overall, the results of this visual analysis fit well with those presented in the previous section, revealing a trend towards reduced variability in pitch over the course of non-serious utterances. It can therefore be concluded that the overall pitch pattern of speech may be manipulated in order to mark non-seriousness as such through reduced pitch range and variation.

3.3.3. Vowel Formants

In addition to measuring and analyzing the pitch values for all speakers, an analysis of their vowel formants was undertaken. In each utterance, an instance of the vowels /i/ and /ɑ/ were located. These vowels were selected in order to approximate the height and width of the vowel space of the speakers, allowing a comparison of any potential changes in the vowel space from serious talk to non-serious talk across speakers, detectable on the basis of shifts in formants. Note that in the varieties of German spoken by the speakers in this study (Berlin-Brandenburgish and Hessian, both considered varieties of Central German) (Wiese, 1996), the vowel /ɑ/ is frequently produced as /a/, such that we can anticipate slightly fronted values for the low back vowel across the board (Wiese, 1996).

In locating an instance of /i/ and /ɑ/ for all speakers in both the serious and non-serious excerpts, several considerations and adjustments were made. First, care was taken to eliminate from consideration any vowel immediately adjacent to (whether following or preceding) the liquids /r/ and /l/ and the approximates /w/ and /j/. This is due to the extreme raising and

lowering effects of these phones on adjacent vowel formants (Ladefoged, 2001), such that any comparison of seriousness with non-seriousness would be unduly influenced by the presence of these consonants. In addition, care was taken to avoid vowels immediately preceding a nasal, such as /n/ or /m/, again, due to the slight formant raising and lowering effects of these phones when following a vowel (Ladefoged, 2001), although in two instances, the inclusion of such vowels was unavoidable (see Excerpts 1 and 8, below). Given that these effects are less extreme than those found with liquids and approximates, it was assumed that the use of these vowels would not have a notable effect on the overall experiment, an assumption which was borne out by additional analysis of Marion's vowels, as described below.

Apart from these general considerations, several speaker-specific adjustments to the selection criteria were necessary in order to allow for the location of vowels which a) fit the above criteria where possible, b) were in a portion of the excerpt that was suitable for acoustic analysis (especially in the case of the non-serious utterances), and c) ideally had been heard by participants in the perception portion of the survey. For Simone (Excerpts 1/2), the only available /ɑ/ vowel in the non-serious excerpt (Excerpt 1) preceded a nasal. Given that the non-serious excerpts are all short in duration and only encompass a single intonational unit, it was necessary to select this vowel. In addition, in order to locate an /i/ vowel in Simone's serious talk, it was necessary to select a vowel in the intonational unit previous to that which was used for the remaining acoustic analyses presented in this chapter. However, the serious utterance containing /i/ had been made immediately previous to the one used for the rest of the analyses, and was also included in the perception survey as part of her serious excerpt (Excerpt 2).

Excerpts 3/4 (Maximilian) both contained /i/ and /ɑ/ vowels which fit all selection criteria outlined above, as did Excerpt 5 (Gerhart). However, for Gerhart's serious utterance (Excerpt 6),

it was necessary again to locate an /ɑ/ vowel in surrounding talk, as there was not one available in the serious excerpt selected for acoustic analysis and use in the survey. Unfortunately, there were no otherwise suitable /ɑ/ vowels in the excerpts included in the survey, so this vowel was selected from serious talk which was located in intervening talk between the serious and non-serious excerpts utilized in the survey.

Marion (Excerpts 7/8) did not happen to produce the vowel /i/ in either the serious or the non-serious excerpts, however, there were numerous instances of /ɪ/. Given that our interest here is in comparing changes in vowel formants relative to the individual speakers, not in exact formant values or mapping of formant data from a large data set, it was possible to compare the changes in /ɪ/ in Marion's speech to changes in /i/ in others' speech. Finally, as briefly noted above, the only instance of /ɑ/ available in her serious utterance (Excerpt 8) was immediately followed by a nasal. This vowel was selected for analysis despite this, on the assumption that the following nasal would have a minimal impact on the results which could be accounted for. Further analysis of vowels in her surrounding speech confirmed this assumption, as described below. Table 3.9. lists the tokens selected for each speaker for vowel formant analysis, with transcript line numbers corresponding to the transcripts presented above for sequential analysis.

Table 3.9.

Vowel Tokens for Formant Analysis (All Speakers)					
Speaker	**Utterance Type**	**Line**	**/i/ Token**	**Line**	**/ɑ/ Token**
Simone (E 1/2)	■ **Non-Serious**	46	*Diamante*	46	*Diamante*
	▲ **Serious**	13	*die*	14	*hab*
Maximilian (E 3/4)	■ **Non-Serious**	39	*schieben*	39	*hat*
	▲ **Serious**	21	*irgendwie*	21	*gemacht*
Gerhart (E 5/6)	■ **Non-Serious**	52	*wie*	54	*macht*
	▲ **Serious**	24	*wie*	42	*hat*
Marion (E 7/8)	■ **Non-Serious**	68	*nicht* (/ɪ/)	68	*das*
	▲ **Serious**	24	*Tisch* (/ɪ/)	27	*Stammtischschild*

After identifying vowels suitable for measuring formants, the relevant data were excerpted as .wav files. For each vowel, the head was located through examination of the wave form in PRAAT. Measurements of F1 and F2 were taken for each excerpt at the head of the vowel at stable points where movement of formants and intensity had plateaued. The settings utilized for measuring formants were uniquely adjusted for each speaker, as summarized in the table below.

Table 3.10.

PRAAT Settings by Speaker				
Speaker	**Maximum Formant (Hz)**	**Number of Formants**	**Window Length(s)**	**Dynamic Range (dB)**
Simone (E 1/2)	5000.00	5.00	0.02	30.00
Maximilian (E 3/4)	4500.00	5.00	0.02	30.00
Gerhart (E 5/6)	5000.00	5.00	0.02	30.00
Marion (E 7/8)	5500.00	5.00	0.02	30.00

On the basis of previous work done in this area, it was anticipated that there would be a trend towards an increase in F1 and F2 in the non-serious excerpts as compared to the serious controls. An increase in F1 corresponds to a lowered vowel, while an increase in F2 corresponds to a more fronted vowel (Ladefoged, 2001). According to Tartter (1980), the combined effect of an increase in both parameters is unrounding, which is a result of the widened lip position and accompanying shortening of the vocal tract observed when a speaker is smiling. On the basis of a large data set consisting of [hVd] vowel tokens, Tartter (1980) and Tartter and Braun (1994) conclude that there is in fact a small but statistically significant tendency for an increase in both F1 and F2 across all vowels when engaging the smile gesture. Using naturally occurring data, Ford and Fox (2010) similarly report an increase in F1 values in non-serious utterances. Auberge and Cathiard (2003; French), however, found a decrease in F1 and an increase in F2 in their experiment using script data. Although in comparison to the laboratory studies, this data set is quite small, it was anticipated that there would be a similar tendency for an increase in F1 and F2 values in the non-serious utterances as compared to the serious utterances.

3.3.3.1. Data: Vowel Formants

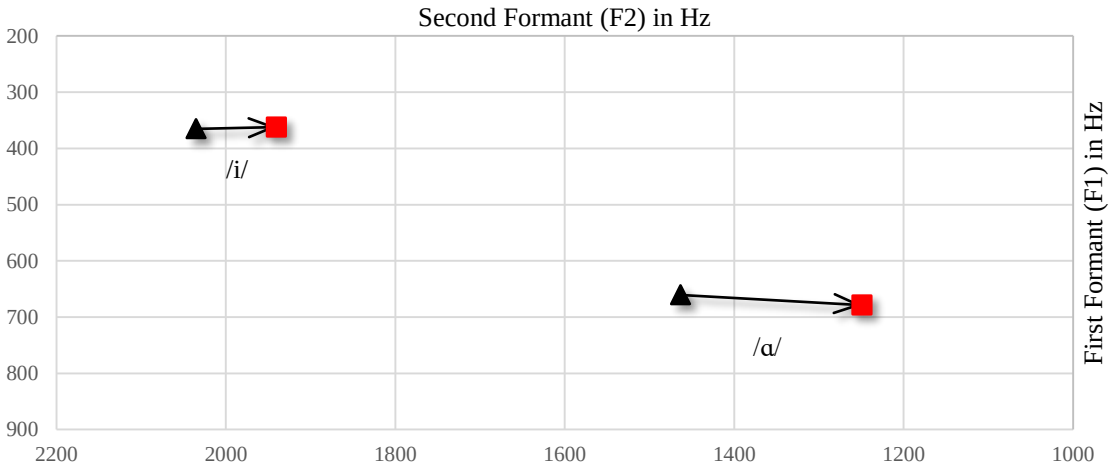
The data for each individual speaker are presented first, followed by an aggregate analysis of the data from all speakers. The vowels /i/ and /ɑ/ are plotted for all speakers in charts which illustrate the vowel space with vowel height (first formant or F1 values) on the vertical axis and vowel frontness (second formant or F2 values) on the horizontal axis. As noted above, an increase in F1 corresponds to a lower vowel, while an increase in F2 corresponds to a more fronted vowel. Whereas traditionally numerical charts are displayed with the lowest values at the bottom and left, the numerical scale of the axes in these charts is reversed from the traditional view in order to illustrate the mouth space, such that the lowest F1 values are at the top of the

vertical axis and the lowest F2 values at the right of the horizontal. This facilitates visualization of the speaker's mouth and relative tongue position, with the mouth and teeth positioned to the left and the throat to the right.

The formant values for Simone (Excerpts 1/2) are presented below in Figure 3.11.

Figure 3.11.

Vowel Formants: Simone (Excerpts 1/2)



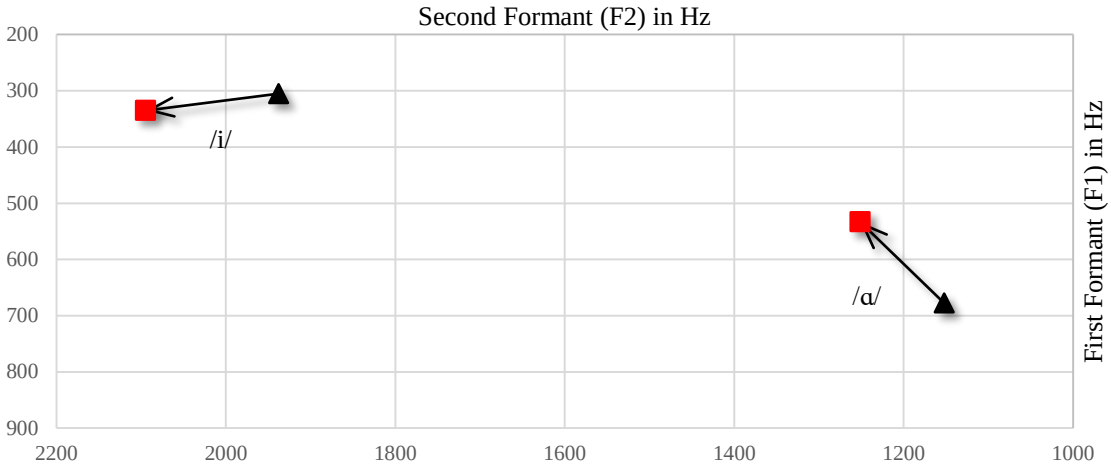
	/i/		/ɑ/	
	F1	**F2**	**F1**	**F2**
▲ **Serious**	365.32	2034.90	660.70	1463.08
■ **Non-Serious**	362.06	1940.10	678.79	1249.07

As we can see in the above chart, Simone's vowels do not show the anticipated increases in F1 and F2 when making a non-serious utterance. In comparison to her serious utterance, Simone's /i/ vowel in her non-serious utterance shows a very small decrease in both F1 and F2, at a ratio of 0.99 and 0.95, respectively. Her /ɑ/ shows a small increase in F1, but a decrease in F2, at a ratio of 1.03 and 0.85, respectively. The changes in F1 are very slight, but the consistent decreases in F2 values, indicating backer vowels in her non-serious utterances, is noteworthy.

Figure 3.12. below contains the formant values for Maximilian (Excerpts 3/4).

Figure 3.12.

Vowel Formants: Maximilian (Excerpts 3/4)



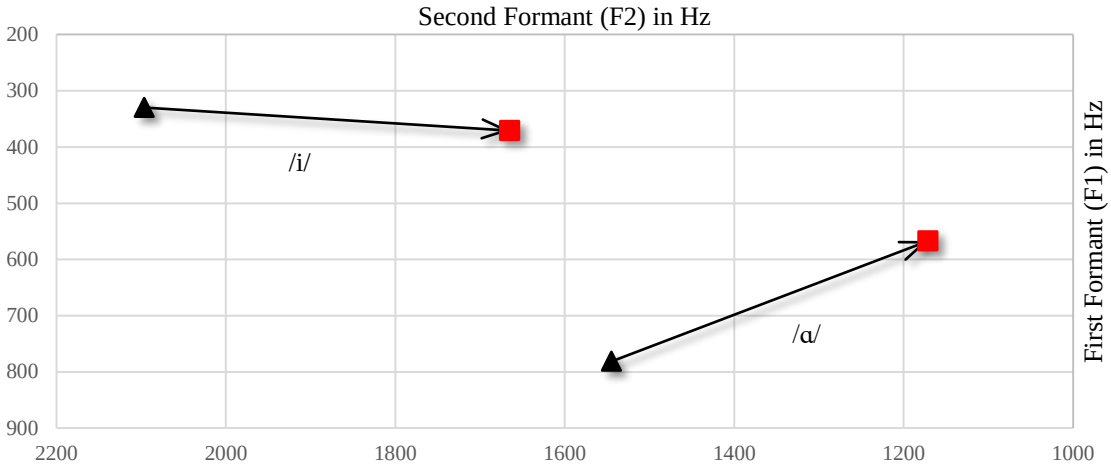
	/i/		/ɑ/	
	F1	**F2**	**F1**	**F2**
▲ **Serious**	305.05	1937.55	677.68	1152.01
■ **Non-Serious**	335.18	2094.66	533.23	1251.16

In contrast to Simone, Maximilian's vowel formant values were moderately higher in his non-serious utterance. Both F1 and F2 were increased in his /i/ vowel, at a ratio of 1.10 and 1.08, respectively. While /ɑ/ had a lower F1value, at a ratio of 0.79, F2 was increased by a ratio of 1.09. These data reveal that Maximilian does have substantial variation in the vowel formants of his non-serious utterance as compared to his serious utterance, indicating that this may be an acoustic variable which he manipulates in order to mark his non-serious utterances. There is a tendency to increase the formant values, as anticipated on the basis of Tartter's (1980) research, although this did not hold for all formants for this speaker. Specifically, the decrease in F1 of the /ɑ/ indicates that the changes in formant values seen in the non-serious utterance may not be fully attributable to the lip spread of the smile gesture.

Formant values for Gerhart (Excerpts 5/6) are summarized in Figure 3.13. below.

Figure 3.13.

Vowel Formants: Gerhart (Excerpts 5/6)



	/i/		/ɑ/	
	F1	**F2**	**F1**	**F2**
▲ **Serious**	329.78	2096.41	781.22	1544.57
■ **Non-Serious**	370.85	1664.72	567.09	1170.10

Similar to the results presented above for Maximilian, we observe here that Gerhart also shows variation in his vowel formants in his non-serious utterance. However, unlike Maximilian, his formant values reveal a tendency to decreased formants in his non-serious utterance as compared to the serious utterance, in contrast to the increase anticipated based on the results from Tartter (1980) described above. Indeed, for the /ɑ/ vowel, Gerhart's F1 and F2 values are both lower in the non-serious utterance than in the serious one, at a ratio of 0.73 and 0.76, respectively. The /i/ vowel also has lower F2 values, at a ratio of 0.79 in comparison to the serious utterance, however, the trend for reduced formant values is broken with the F1 measure. Here, we see that the F1 of /i/ in Gerhart's non-serious utterance is higher than in the serious utterance, at a ratio of 1.12.

Note that although these values run counter to the expectation that vowel formants will be

increased for non-serious utterances due to the unrounding associated with the physiological effects of the smile gesture, it is entirely possible to make a non-serious utterance without smiling. While the utterance itself is indeed analyzablely treated as non-serious by the interlocutors, as demonstrated in the sequential analyses presented above, we observe in the video data stream that Gerhart does not appear to engage the smile gesture until *after* the utterance is completed and he has begun to invite laughter from Donnie. The progression of Gerhart's lip gestures over the course of the non-serious utterance (Excerpt 5) can be seen in Figure 3.14., below.

Figure 3.14.

Progression of Smile Gesture: Gerhart (Excerpt 5)

In the above series of screenshots, we observe Gerhart's lip position prior to the non-serious utterance (A), during the /i/ vowel (B), during the /ɑ/ vowel (C), and following the utterance as

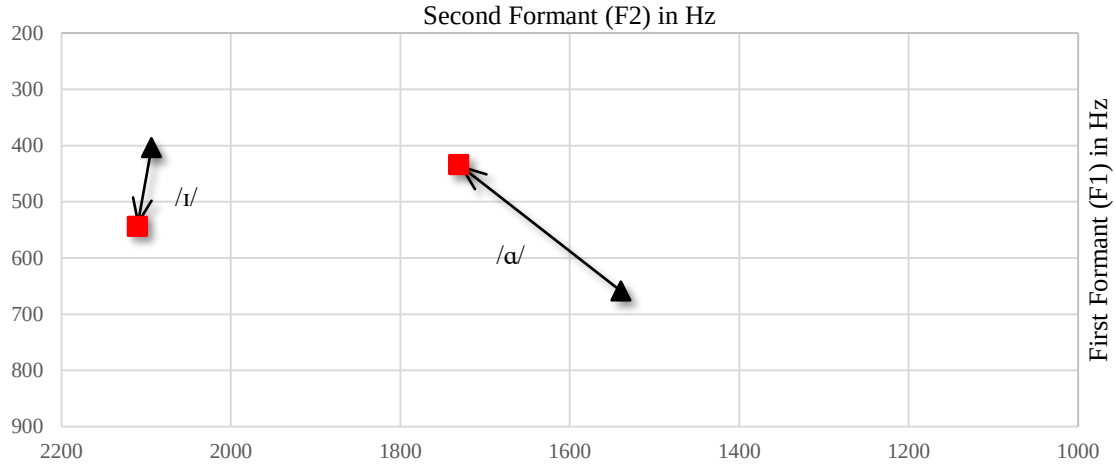
he begins to laugh (D). It is clear that he does not begin to make the smile gesture until *after* the non-serious utterance, suggesting that the mere presence or absence of the smile gesture alone cannot account for the variations in formant values observed in non-serious utterances. Rather, it may be the case that speakers could produce an utterance with formant values which differ from their normal surrounding speech, whether by lower or higher, purely as a means of marking non-seriousness, regardless of the presence or absence of the smile gesture.

On the basis of the formant values presented above, despite the lack of smile gesture, I conclude that Gerhart does likely manipulate his vowel formants as a means of marking his utterance as non-serious. However, we see that in this case, manipulation of vowel formants is reflected in a trend for marked decrease rather than increase.

Figure 3.15., below, illustrates the formant values for Marion (Excerpts 7/8).

Figure 3.15.

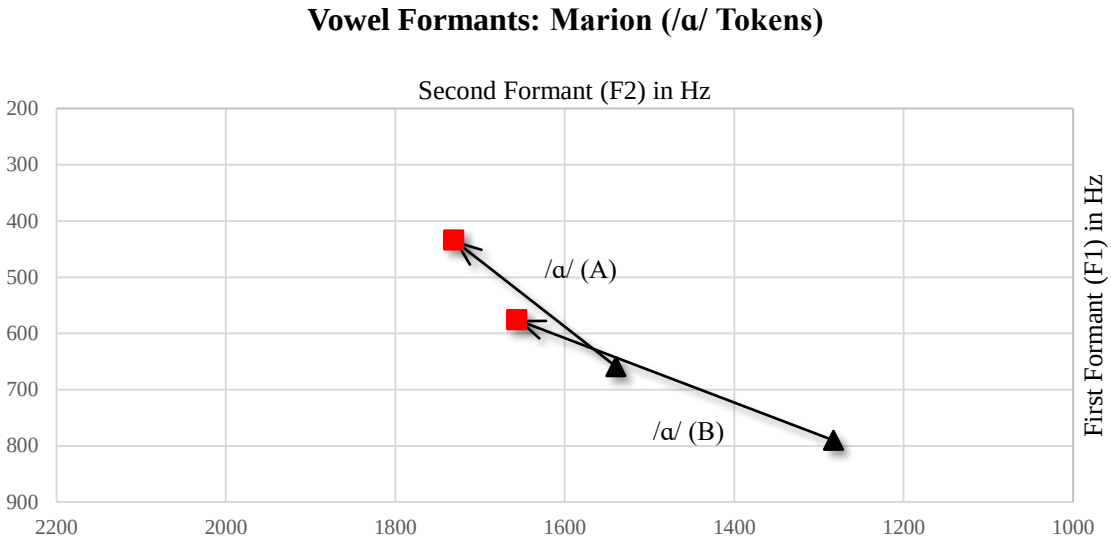
Vowel Formants: Marion (Excerpts 7/8)



	/ɪ/		/ɑ/	
	F1	**F2**	**F1**	**F2**
▲ **Serious**	403.70	2093.72	659.05	1539.34
■ **Non-Serious**	543.81	2110.38	433.92	1730.94

As noted above, it was not possible to locate an instance of /i/ in the utterances selected from Marion's speech for further analysis. In order to analyze the formant changes from seriousness to non-seriousness in a high front vowel from this speaker, instances of /ɪ/ were selected instead. The results indicate that there was a substantial increase in the F1 value of /ɪ/ in her non-serious utterance as compared to the serious utterance, at a ratio of 1.35, while there was only a negligible increase in F2, at a ratio of 1.01. The F1 value of /ɑ/ in Marion's non-serious utterance decreased in comparison to the serious one, at a ratio of 0.66, whereas the F2 value increased at a ratio of 1.12. These results indicate that vowel formants are likely an acoustic variable which Marion manipulates in order to mark an utterance as non-serious, albeit in an interesting pattern: while the F1 of /ɪ/ and the F2 of /ɑ/ both increased, the F2 of /ɪ/ remained nearly static and the F1 of /ɑ/ decreased. When measuring further instances of Marion's /ɑ/ vowel from *da* (Excerpt 7, line 68; non-serious) and *Wappen* (Excerpt 8, line 29; serious), regardless of the presence of a subsequent nasal consonant, the pattern of a decrease in F1 (raising) and an increase in F2 (fronting) in non-serious utterances held, as can be seen in Figure 3.16. (where the vowels in Group A are the same as those charted above in Figure 3.15.).

Figure 3.16.



	/ɑ/ (A)		/ɑ/ (B)	
	F1	**F2**	**F1**	**F2**
▲ **Serious**	659.05	1539.34	790.23	1282.39
■ **Non-Serious**	433.92	1730.94	575.63	1656.68

Here, we see that, regardless of the presence of a nasal consonant immediately following the /ɑ/ vowel in the non-serious utterance in Group A, the pattern of a decrease in F1 and an increase in F2 as compared to the serious utterance holds. It appears then that Marion manipulates the formants of her vowels in her non-serious utterance such that they differ in a consistent manner from the serious utterance, although this may not be attributable solely to lip spread and accompanying unrounding, as also noted above in Maximilian's speech.

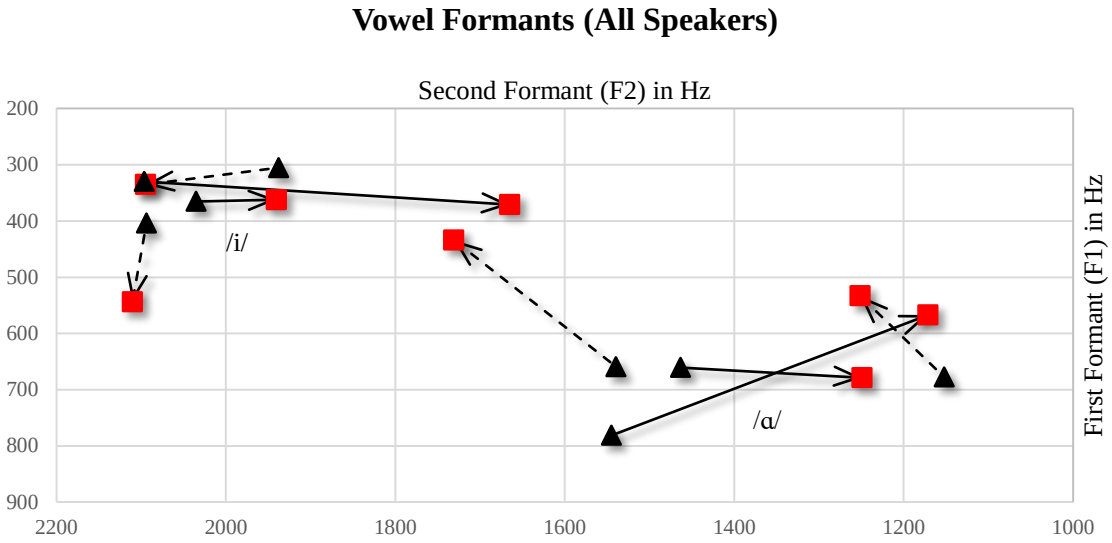
3.3.3.2. Results: Vowel Formants

The data above reveal that most speakers in this study tended to have substantially differing formant values in their non-serious utterances as compared to the serious utterances. A possible exception was Simone; the changes in her vowel F1 formant values were very small, such that they were potentially imperceptible to her interlocutor. It may be the case that she was

not employing changes in F1 relative to surrounding talk as a cue for non-seriousness, or that the changes were simply very subtle in this particular instance.

Considering the formant values of all speakers in aggregate, several interesting patterns emerge which lend support to the conclusion that speakers may manipulate vowel formant values of non-serious utterances. On the basis of previous research, it had been anticipated that the lip spread associated with the smile gesture would result in unrounding of the vowels, i.e., in an increase in both F1 and F2, indicating raising and fronting of the vowels (Tartter, 1980). However, a comparison of the changes in vowel formant values seen in all speakers, as can be seen in Figure 3.17. (below), clearly demonstrates that this anticipated pattern of consistent increases in F1 and F2 were not found to hold in these data.

Figure 3.17.



Speaker	Utterance Type	/i/		/ɑ/	
		F1	F2	F1	F2
Simone (E 1/2)	▲ Serious	365.32	2034.90	660.70	1463.08
	■ Non-Serious	362.06	1940.10	678.79	1249.07
Maximilian (E 3/4)	▲ Serious	305.05	1937.55	677.68	1152.01
	■ Non-Serious	335.18	2094.66	533.23	1251.16
Gerhart (E 5/6)	▲ Serious	329.78	2096.41	781.22	1544.57
	■ Non-Serious	370.85	1664.72	567.09	1170.10
Marion (E 7/8)	▲ Serious	403.70	2093.72	659.05	1539.34
	■ Non-Serious	543.81	2110.38	433.92	1730.94
Region	Berlin-Brandenburgish	→	Hessian	- - -→	

If changes in formant values could be solely attributed to unrounding caused by the lip spread of the smile gesture (Tartter, 1980; Tartter & Braun, 1994), then we would anticipate that all of the arrows in the above figure would point more or less in the same direction, namely, trending down and to the left. This is simply not the case here.

The data presented above, therefore, do not support the conclusion that such manipulation

is necessarily the direct result of unrounding. Rather, the data reveal three trends in variation of formants from a baseline value to indicate a shift in [+/- humor] status: 1) such variation varies from speaker to speaker, while 2) such variation follows a discernible pattern on the intraspeaker level and 3) may correlate with native variety of German spoken. To explicate the first of these trends, we again refer to Figure 3.17., above. Note that the vowels /i/ and /ɑ/ as found in the serious utterances (indicated in the chart by the black triangles, ▲) tend to bunch together in the vowel spaces associated with these vowels on the basis of larger population studies (Kohler, 2008). This result was expected, as we would generally expect that, in a given sample population, the vowels of said population will tend to fall within the mean values of the larger population from which the sample was taken.

If we then consider the direction of formant shift from each serious utterance to the associated non-serious utterance from the same speaker, it can be seen that there is no clear trend. In other words, in comparison to the serious utterances, some vowels in the non-serious utterances are higher, some lower, some fronter, some backer, with no discernible trend immediately obvious. Indeed, the vowels of the non-serious utterances (indicated by the red squares, ■) appear to be scattered across the plot in an almost random fashion in comparison to the serious utterances. This leads to our first conclusion, i.e., the way in which vowel formants in non-serious utterances vary from their baseline serious counterparts differs from speaker to speaker.

That being the case, if we consider again Figures 3.11., 3.12., 3.13., and 3.15., we observe a further trend which may not have been immediately obvious on initial examination of the individual data sets: There is a tendency on the intra-speaker level for speakers to manipulate their individual vowel formants in a recognizable pattern unique to that speaker, especially in

terms of the relative frontness of the vowels in their non-serious utterances as measured by changes in F2. Specifically, both Simone's (Excerpts 1/2) and Gerhart's (Excerpts 5/6) vowels in their non-serious utterances are backer than in the serious utterances, while both Maximilian's (Excerpts 3/4) and Marion's (Excerpts 7/8) non-serious vowels are fronter. The consistency of the movement of the vowels within speakers is striking indeed, especially in light of the additional analysis of Marion's vowels, shown in Figure 3.16., which revealed that her /ɑ/ vowel was consistently higher and fronter in two non-serious utterances in comparison to her serious utterances. This result leads to our second conclusion, that is, while the way in which speakers manipulate their vowels in non-serious utterances from their individual baseline varies by speaker, such variation is strikingly consistent within a given speaker's speech, especially in terms of changes in F2 values.

The final key observation is that the individual patterns of shifts in formant values outlined in the previous paragraph correlate neatly with regional varieties of German. That is, as described above, Simone and Gerhart, both speakers of Berlin-Brandenburgish (indicated above in Figure 3.17. by solid arrows), have backer vowels in their non-serious utterances, while Maximilian and Marion, both speakers of Hessian (dashed arrows), have fronter vowels. Absent additional data, it is impossible to say whether this correlation is indeed caused by regional variation. It is nonetheless an interesting feature of the data which is worth noting and is in need of further exploration.

On the basis of these data, it can be concluded that manipulation of vowel formants from a baseline norm can be used as a marker of non-seriousness. Such manipulation varies from speaker to speaker, that is, there is no one way to mark non-seriousness with vowel formants, and in fact may be very subtle or even absent in the speech of some speakers, such as is the case

with Simone's (Excerpts 1/2) F1 values. However, such manipulation also appears to be consistent on the intra-speaker level, that is, each speaker manipulates their vowel formants in a consistent pattern unique to that speaker in order to mark a given utterance as non-serious. Furthermore, there is a potential correlation of this pattern with regional variation. This evidence suggests that manipulation of vowel formants from a normal baseline unique to individual speakers is indeed likely a member of the constellation of features which may be manipulated by speakers in order to mark an utterance as non-serious.

3.3.4. Intensity

A further acoustic feature measured for this study was intensity. Using PRAAT, the intensity of each utterance was measured to determine the maximum, minimum, and average values. Maxima and minima were first automatically detected in the audio signal using the software. The waveforms were then visually and acoustically inspected to identify any potential measurement errors, such as false readings of background noise. No such errors were identified. Similarly, the average values were calculated by selecting the relevant portion of the audio signal and allowing the software to calculate the average over the course of the selected portion of the audio signal only, minimizing the influence of background noise in non-selected portions of the audio signal on the resultant value.

In addition to measuring absolute changes in intensity, it is also necessary to consider scaled intensity values. Unlike pitch and vowel formants, intensity, or volume, can easily be influenced by factors purely exogenous to the acoustic properties of the speech stream, such as microphone placement or a slight movement of the head towards or away from the microphone between utterances. Scaling raw intensity values allows for better comparison of variation across speakers and recording contexts by measuring changes relative to a uniform average. In this

case, the intensity was scaled across all excerpts at 70 dB. The measurements of maximum, minimum, and average were then taken again for the normalized excerpts following the same procedure described above.

On the basis of previous research in the field, it was anticipated that intensity was likely to be a component of the constellation of features which may be manipulated in order to mark an utterance as non-serious. However, whether such manipulation should have resulted in increased or decreased intensity values was unclear, as previous research had described varying results. Most studies documented an increase in intensity. Amongst these, Anolli et al. (2000; Italian), Auberge and Cathiard (2003; French), Rockwell (2000), and Tartter (1980) based their results on reading script data, while Purandare and Litman (2006) analyzed television data and Archakis, Giakoumelou, Papazachariou, and Tsakona (2010; Greek) considered conversations held in a laboratory setting. However, Cheang and Pell (2008, 2009, 2011) reported both increases and decreases in intensity for English and Cantonese reading script data, depending on whether the reading prompt was humorous or sarcastic. The two studies which utilized naturally occurring data reported conflicting results, with Flamson et al. (2011) finding an increase in intensity and Hidalgo Navarro (2011; Spanish) a decrease.

A handful of studies, however, reported no significant change in intensity. Attardo et al. (2011), Pickering et al. (2009), and Voyer and Techentin (2010) based these results on reading script data, while Bryant and Fox Tree (2005) had similar results with radio data. One study using conversational data collected in a laboratory, Bryant (2010), also found no significant changes in intensity, while also finding a decrease in variation in intensity.

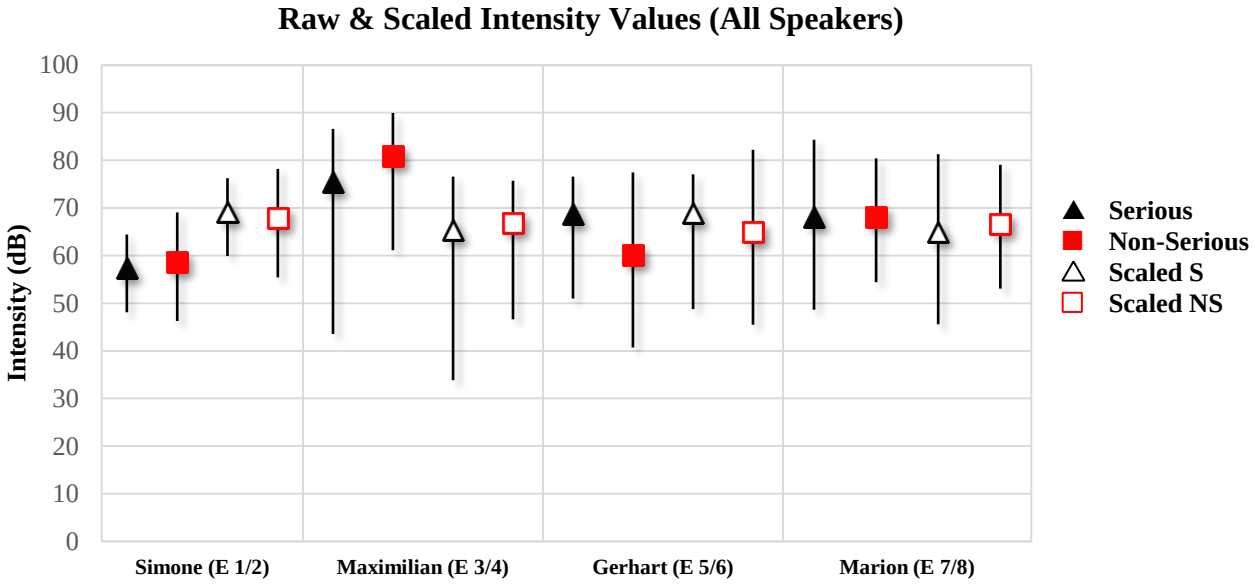
In light of these studies, it was anticipated that speakers of German in naturally occurring conversations were likely to manipulate intensity in the service of marking an utterance as non-

serious. The data presented below reveal that this is likely the case. However, similar to what was found in the review of the extant literature described above, there was no clear trend across all speakers towards an increase or a decrease in intensity, although changes in variation of intensity present an interesting case for further consideration.

3.3.4.1. Data: Intensity

The maximum, minimum, and average raw and scaled intensity values for all speakers are presented in Figure 3.18., below.

Figure 3.18.

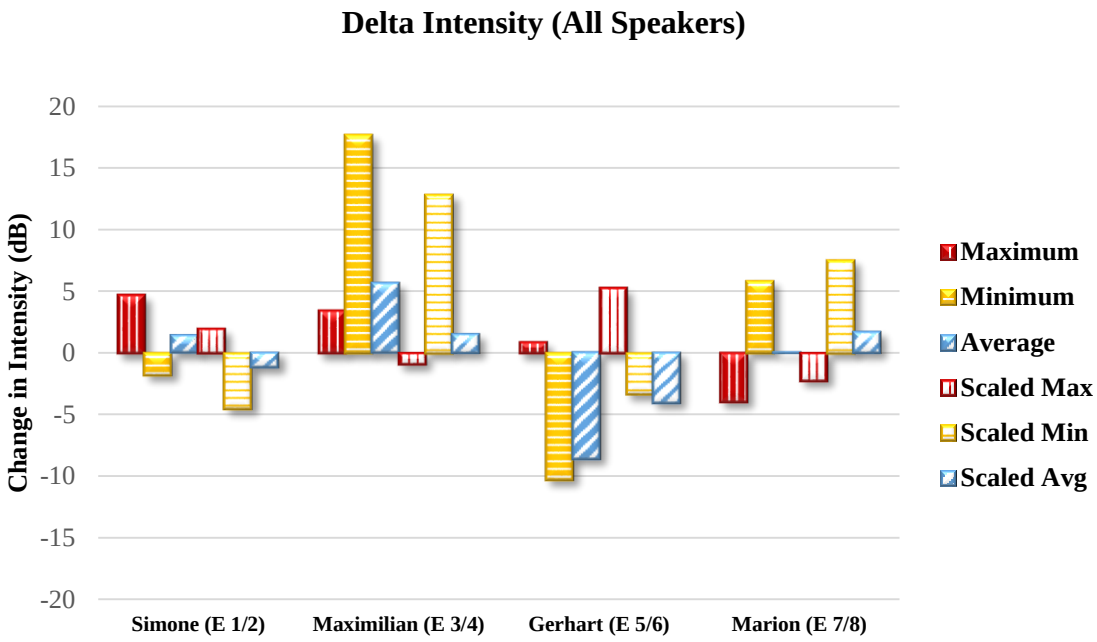


Speaker	Maximum				Minimum				Average			
	▲	■	△	□	▲	■	△	□	▲	■	△	□
Simone (E 1/2)	64.4	69.0	76.3	78.2	48.1	46.3	59.9	55.4	57.3	58.6	69.1	67.9
Maximilian (E 3/4)	86.6	89.9	76.6	75.7	43.5	61.1	33.8	46.6	75.4	80.9	65.4	66.9
Gerhart (E 5/6)	76.6	77.5	77.1	82.2	50.9	40.7	48.8	45.5	68.7	60.1	68.8	64.8
Marion (E 7/8)	84.3	80.4	81.3	79.1	48.6	54.4	45.6	53.1	68.2	68.1	65.0	66.7

As can be seen in the above figure, while there are changes in intensity values in non-serious utterances, a consistent pattern is difficult to identify. Simone, Maximilian, and Gerhart all have

higher maximum intensity for their non-serious utterances, yet this result does not hold for the scaled values. Minimum intensity, on the other hand, is higher for two of the speakers (Maximilian and Marion) and lower for the other two (Simone and Gerhart), which does hold for the scaled values. Similarly, average intensity is higher for two speakers (Simone and Maximilian), lower for one (Gerhart), and nearly the same for the final speaker (Marion). A summary of delta values can be viewed in Figure 3.19., below.

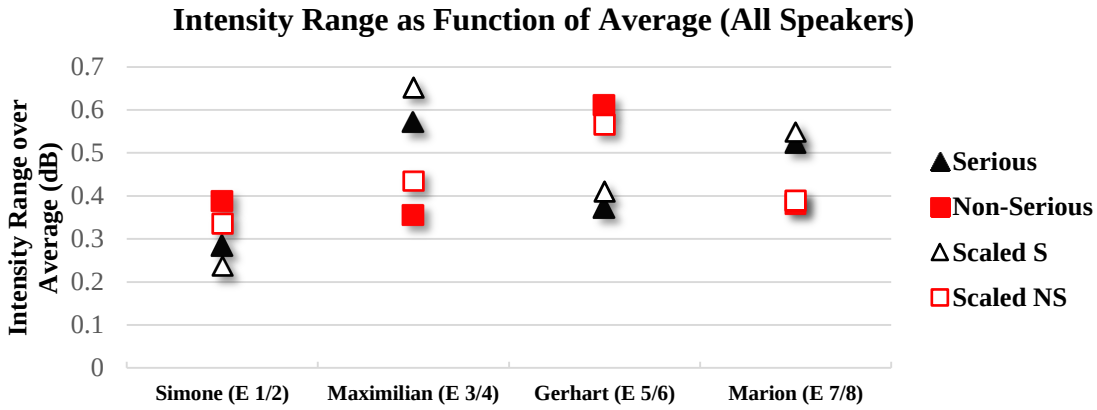
Figure 3.19.



Again, as described above, while there are certainly small changes in intensity, any clear pattern is difficult to discern, whether considering the raw intensity values or the scaled data.

Considering the range of the intensity values, we observe that two speakers (Simone and Gerhart) have an increased intensity range, while Maximilian and Marion have a decreased intensity range. Figure 3.20., below, charts the intensity range of all speakers as a function of the average.

Figure 3.20.



Here, we again observe that the participants are evenly split. Simone and Gerhart both have increased variation in intensity over the course of their non-serious utterances, while Maximilian and Marion both have decreased variation, both for raw intensity measures and for the scaled values.

3.3.4.2. Results: Intensity

As described above, previous research has claimed that changes in intensity are a feature of non-serious utterances or jokes. However, the literature has not been unanimous as far as whether there is a trend for increases or decreases in intensity values. Given that, it is not surprising that the data presented above show clear differences in the intensity values for all speakers in their non-serious utterances as compared to their serious ones, while at the same time, there is no clear, discernible pattern to the changes.

One interesting observation is that there is again a neat correlation with regional varieties. Both Simone and Gerhart, speakers from Berlin-Brandenburg, have increased variation in intensity over the course of the utterance, while Maximilian and Marion, speakers of Hessian, have decreased variation in intensity. This result holds for both raw intensity values and the scaled data. Given that the results of the formant analysis presented above suggested a similar correlation of manner of marking non-seriousness with regional variety, further regional and

dialectal analysis to determine whether these trends hold or are quirks of the small data set is worth pursuing.

While there is no clear trend for all speakers in how intensity in non-serious utterances varies from baseline serious values, the data presented above show that all non-serious utterances do show some variation from baseline values. The evidence for intensity as a component of the constellation of acoustic features which are manipulated in order to mark non-serious utterances is indeed weaker than for other acoustic features presented here due to this lack of a clear trend or pattern. However, it remains clear that there are distinct changes in intensity, and that these changes vary substantially by individual speaker (although there is some evidence to suggest there may be a pattern for changes in variation in intensity which correlates with regional variety of German spoken), a result which is consistent with the evidence presented for changes in vowel formants.

3.3.5. Beats

In addition to pitch, vowel formants, and intensity, a phenomenon known as beats was analyzed. Beats refer to rhythmic patterns of stressed syllables in speech (Couper-Kuhlen, 2009). According to Couper-Kuhlen, stressed syllables are those syllables which have prosodic prominence created by a confluence of features primarily consisting of increased length, pitch, and intensity relative to surrounding unstressed syllables. Building on the research tradition of rhythm in speech, she demonstrates that participants in conversation orient to changes in beats of stressed syllables, such that variation in stress rhythm is not simply a randomly distributed artifact of human speech production, but is in fact an interactionally relevant feature of conversation.

Couper-Kuhlen's (1993, 2009) research is based on identification of beats in naturally

occurring conversation using an impressionistic auditory method. That is, beats are identified by listening to the data and annotating the transcript at points of prosodic prominence. The intervals between the beats are then calculated by measuring the time between vowel onset of each stressed syllable as viewed in a spectrogram. As she demonstrates, speakers orient to points of anisochronous changes in the ongoing rhythmic pattern.

While beats, as defined in this way, certainly warrant further investigation, the method used by Couper-Kuhlen is subject to the perceptual limitations of the researcher. This cannot be fully overcome by combining results from independent analyses of spates of talk done by multiple researchers, as done by Couper-Kuhlen, nor is it a necessarily a practical approach. In an attempt to rectify this situation, Ogden at the 17th Working Conference on Research on Conversation and Discourse (17. Arbeitstagung zur Gesprächsforschung) in Mannheim, Germany in the spring of 2013, encouraged practitioners to seek out a unified definition of beats which could be quantified via easily replicable means. While such means do not yet exist to my knowledge, one suggestion made by Ogden (2013) was to utilize a set of PRAAT and PERL scripts published online by Fred Cummins (n.d.) to quantify beats in a given spate of talk on the basis of an analysis of local variations in intensity.

Using this pair of scripts as a starting point, beats were defined as rhythmic local peaks in intensity over the course of an utterance for the purposes of this analysis. While changes in pitch and syllable length are also components of stress, as noted above, the decision was made to focus on intensity for several reasons. First, given that the scripts offered by Cummins analyze intensity only, it seemed appropriate to base my analysis on this aspect of stress as well. Further, as machine identification of beats is in its nascent stages, selecting one feature of stress allows for a clean initial analysis. Future attempts should incorporate consideration of pitch and

syllable length as the technology to do so develops. Finally, local variation in pitch had already been considered as a separate component of non-seriousness, as presented earlier in the chapter.

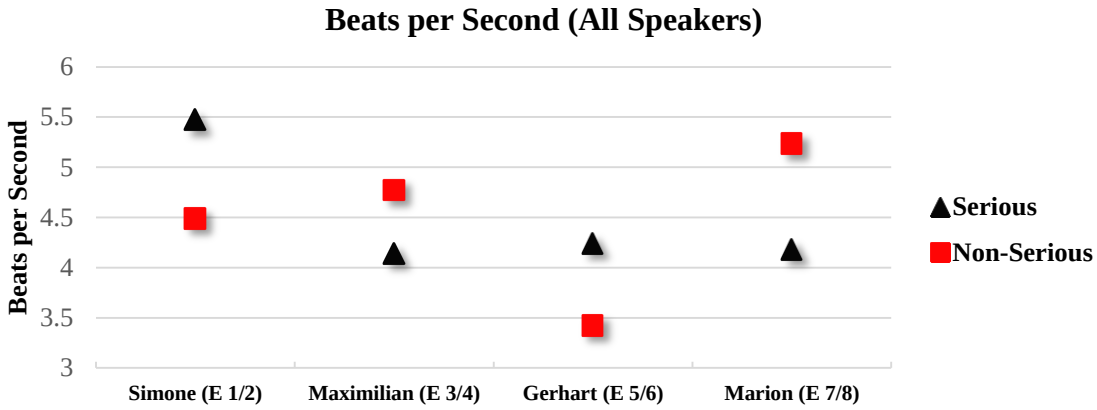
With the assistance of a colleague, Nicholas Williams, a new script was written to replace the PRAAT and PERL method available from Cummins (n.d.). This script, which can be viewed in Appendix 3.2., runs solely in PRAAT. It first measures the maximum, minimum, and average intensity of the selected excerpt, then scales the results at an average of 70 dB and scans the utterance for local peaks in intensity. These local peaks in intensity are then counted as beats by the program, which can also be viewed in a text grid.

Each serious and non-serious utterance was first analyzed using the PRAAT script found in Appendix 3.2. The resultant text grids were then reviewed in order to eliminate from consideration any beats which were potentially associated with background noise; none were found. The number of beats located within each excerpt were counted and taken over the duration in order to calculate the average rate of beats per second for each utterance. The inverse, duration in seconds over beats, was used to calculate the average interval between beats.

3.3.5.1. Data: Beats

Taking the number of beats counted in each excerpt over the length of the excerpt in seconds (beats per second) allows for comparison of the frequency of beats across speakers. Figure 3.21., below, contains a summary of the number of beats per second for all speakers.

Figure 3.21.



Speaker	▲	■	▲ Average Interval	■ Average Interval
Simone (E 1/2)	5.47	4.49	0.183 s	0.223 s
Maximilian (E 3/4)	4.14	4.78	0.241 s	0.209 s
Gerhart (E 5/6)	4.24	3.42	0.236 s	0.292 s
Marion (E 7/8)	4.18	5.23	0.239 s	0.191 s

As can be seen in the above chart, both Maximilian and Marion have more beats per second in their non-serious utterances, while Simone and Gerhart have fewer beats on the same measure. The average beat interval, and therefore the number of beats per second, is very similar for the serious utterances of Maximilian, Gerhart, and Marion. Simone's non-serious utterance is actually quite similar to the other speakers' serious utterances on this measure, which is a somewhat unusual result.

3.3.5.2. Results: Beats

The above data reveal that two of the speakers, Maximilian and Marion, have increased beats per second in their non-serious utterances, while the other speakers (Simone and Gerhart) have decreased beats per second. These results again show a difference based on regional variation, as we have seen previously with vowel formants and intensity. Given that the pattern of beats is very similar for Maximilian, Gerhart, and Marion for their serious utterances, it may

be the case that changes in a standard rhythm of beats in unmarked, unemphatic serious utterances can be used as a means of setting an utterance apart from the rest of the talk. Further research is necessary in order to determine what, if any, role beats play in the marking of non-seriousness in everyday conversation.

3.3.6. Speech Rate

The rate of speech can be measured in a variety of ways, however, the most common methods consider the number of syllables produced in relation to duration. To my knowledge, all studies on humor which measure speech rate have used this method, presenting their results either as a rate of syllables per second, or as average length of time spent uttering each syllable (i.e., seconds per syllable). The results have varied greatly, with no trends suggested by the literature. Archakis et al. (2010; Greek; conversation in laboratory), Bryant (2010; conversation in laboratory), and Purandare and Litman (2006; television) found increased speech rate, while Cheang and Pell (2008, 2009, 2011; Cantonese and English), Tartter (1980), and Tartter & Braun (1994), all using reading script data, found decreased speech rate for humorous utterances. Differentiating between irony and sarcasm, Anolli et al. (2000; Italian), found both increased and decreased speech rate. Most studies, however, found no significant change in the speech rate of humorous utterances, including research from Attardo and Pickering (2011; reading script), Attardo et al. (2011; reading script), Bryant and Fox Tree (2005; radio), Flamson et al. (2011; naturally occurring conversation), and Pickering et al. (2009; interview data).

On the basis of these studies, it was not anticipated that speech rate would be a productive feature of non-seriousness. For the purposes of this experiment, speech rate was defined as the number of syllables per second. The number of syllables for each utterance was determined on the basis of the speech produced (i.e., prescriptive syllabification rules were not

considered, only what was actually produced by the speakers). Tables 3.22.-25., below, contain the full text of each utterance analyzed in this study in both Standard German and as produced by the speaker, a syllabic gloss with the symbol " · " between each syllable counted, a total count of the syllables, the duration of each utterance (measured in seconds using PRAAT), and notes on how instances of ambiguous syllabification were resolved.

Table 3.22.

Syllables for Speech Rate Analysis: Simone (Excerpts 1/2)				
Excerpt	**Utterance**			
1 ■	**Standardized High German**	Es war bestimmt ein Diamant drin.		
	As Produced	es war bestimmt en dia↑mAnte drinne,		
	Syllabification	es · war · be · stimmt · en · dia · man · te · drin · ne		
	Syllables	10	**Duration (s)**	1.56
	Notes	• A schwa sound is clearly audible at the end of *Diamante*; it is unclear if this is an archaic dialectal feature, or if a better transcription would be *Diamant da drinne*; both options result in the same syllable count. • The first syllable of *Diamante* is produced as one syllable rather than two, with a single sonority peak.		
2 ▲	**Standardized German**	Ich habe immer versucht, eine Versteinerung zu finden.		
	As Produced	↑ich hab immer versucht ne verstEInerung zu finden,		
	Syllabification	ich · hab · imm · er · ver · sucht · ne ver · stein · er · ung · zu · fin · den		
	Syllables	14	**Duration (s)**	2.01
	Notes	• A second sonority peak is audible on *finden*, with a syllabic /n/ producing a second syllable.		

Table 3.23.

Syllables for Speech Rate Analysis: Maximilian (Excerpts 3/4)				
Excerpt	**Utterance**			
3 ■	**Standardized High German**	Ja, noch mehr Ball hin- und herschieben, das hat uns gefehlt.		
	As Produced	[ja:]noch- noch mehr: bAll hin und herschieben, ↑dAs hat uns gefehlt;		
	Syllabification	ja · noch · noch · mehr · ball · hin · und · her · schieb · en · das · hat · uns · ge · fehlt		
	Syllables	15	**Duration (s)**	3.56
	Notes	• Both productions of *noch* counted as one syllable each, although the first is slightly cut off. • A second sonority peak is audible on *schieben*, with a syllabic /n/ producing a second syllable.		
4 ▲	**Standardized German**	Und dann haben die es wirklich so bis zum Erbrechen gemacht bei der EM irgendwie, achtzig Prozent Ballbesitz und zwei Torschüsse.		
	As Produced	und dann ham die es (0.1) wirklich so bis zum erbrE:chen gemacht bei der ee em [irgend]wie, achtzig prozent bA:llbesitz, und (0.3) zwEI: torschüsse,		
	Syllabification	und · dann · ham · die · es · wirk · lich · so · bis · zum · er · brech · en · ge · macht · bei · der · ee · em · ir · gend · wie · acht · zig · pro · zent · ball · be · sitz · und · zwei · tor · schüss · e		
	Syllables	34	**Duration (s)**	7.00
	Notes	• The verb *haben* is fully collapsed to one syllable, produced here as /hɑm/ with a single sonority peak.		

Table 3.24.

Syllables for Speech Rate Analysis: Gerhart (Excerpts 5/6)				
Excerpt	**Utterance**			
5 ■	**Standardized High German**	Wie uncool. Wer macht denn das, bitte?		
	As Produced	wie uncool, (0.2) wer macht denn ↑dAs bitte.		
	Syllabification	wie · un · cool · wer · macht · denn · das · bi · tte		
	Syllables	9	**Duration (s)**	1.46
	Notes	• Well articulated with clear syllabification.		
6 ▲	**Standardized German**	Er ist wie ich, ohne zu reden.		
	As Produced	[er ist] wie ↑ich ohne zu reden;		
	Syllabification	er · ist · wie · ich · oh · ne · zu · red · en		
	Syllables	9	**Duration (s)**	1.18
	Notes	• A second sonority peak is audible on *reden*, with a syllabic /n/ producing a second syllable.		

Table 3.25.

Syllables for Speech Rate Analysis: Marion (Excerpts 7/8)				
Excerpt	**Utterance**			
7 ■	**Standardized High German**	Ne, das ist ganz weit weg, da komme ich nicht hin.		
	As Produced	↑nee das ist gAnz wEIt wEg;=da komm ich nicht hin.		
	Syllabification	nee · das · ist · ganz · weit · weg · da · komm · ich · nicht · hin		
	Syllables	11	**Duration (s)**	1.72
	Notes	• There is no audible schwa on the end of *komm*, forming only one syllable.		
8 ▲	**Standardized German**	Ich hätte auch gern so ein Schild für den Tisch. Weißt du, mit so einem Stammtischschild.		
	As Produced	ich hätt auch [gern] so en schI:ld für den tisch; (0.3) weißte mit so nem stAmmtischschild,		
	Syllabification	ich · hätt · auch · gern · so · en · schild · für · den · tisch · weiß · te · mit · so · nem · stamm · tisch · schild		
	Syllables	18	**Duration (s)**	3.35
	Notes	• There is no audible schwa on the end of *hätt*, forming only one syllable. • Both articles (*en* / *ein*, *nem* / *einem*) are produced as single syllables.		

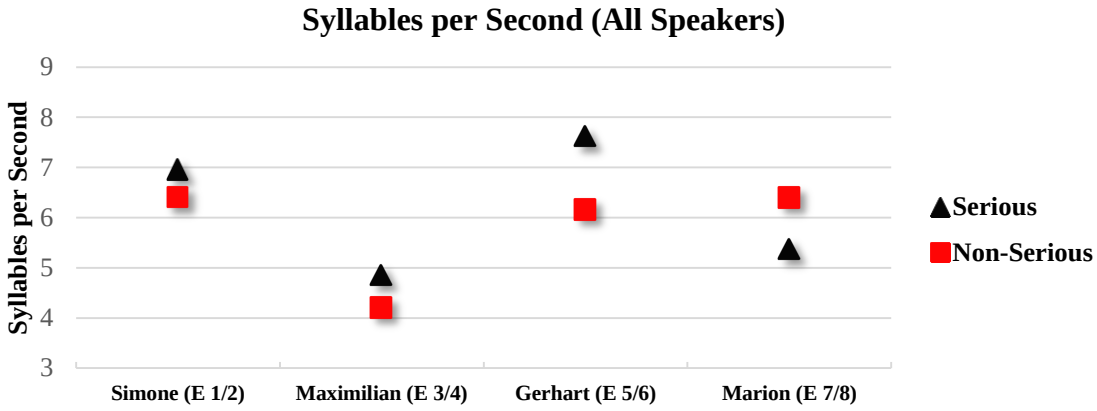
On the basis of the above syllable counts, the speech rate for each speaker was determined by taking the number of syllables over the duration of the utterance in seconds. The data presented below suggest a trend towards decreased speech rate in non-serious utterances as compared to serious ones, a result which indicates that speech rate should be considered in future testing of a larger sample set.

3.3.6.1. Data: Speech Rate

Taking the number of syllables in each utterance over the duration of said utterance gives

the number of syllables produced on average per second, which may then be compared across speakers. Figure 3.26., below, contains a summary of the results of the speech rate analysis.

Figure 3.26.



Speaker	▲	■	Δ ■ → ▲	Ratio ■ / ▲
Simone (E 1/2)	6.97	6.41	-0.55	0.92
Maximilian (E 3/4)	4.86	4.21	-0.64	0.87
Gerhart (E 5/6)	7.63	6.16	-1.46	0.81
Marion (E 7/8)	5.37	6.40	1.02	1.19

As mentioned above, the data reveal a trend towards reduced speech rate (i.e., fewer syllables uttered per second) in non-serious utterances as compared to serious ones from the same speaker. Simone, Maximilian, and Gerhart all produced fewer syllables per second during their non-serious utterances.

3.3.6.2. Results: Speech Rate

As described above, the lack of unity in defining speech rate complicates comparisons to previous studies on speech rate in humor. However, there appears to be a trend in these data towards a reduced rate of speech, a result which is not at odds with previous findings. Further research on a larger data set is necessary to determine if this result stands or is an artifact of the small sample size.

3.4. Conclusion

The first half of this chapter offered a sequential analysis of the excerpts utilized for this study. On the basis of that analysis, it was concluded that all excerpts selected for use in the study were solid examples of non-seriousness based on the CA definition, with evidence of similar sequential features fitted to the definition of non-seriousness derived by Sacks (1972). Specifically, all utterances were treated by both participants in the conversation as non-serious with laughter (with no signs of trouble and/or repair of the laughter), there was no serious uptake of the utterance to which the laughter was responsive (with no signs of repair of this lack of serious uptake), and there was evidence of a jointly managed return to seriousness subsequent to laughter.

All excerpts were then analyzed acoustically using PRAAT, and the results compared to serious utterances from immediately contiguous speech from the same speaker. It was anticipated that a constellation of acoustic features are manipulated by speakers in order to mark an utterance as non-serious. In other words, there would be no one unified "smile voice" consistently marked by a single feature or set of features. Rather, as argued in the introduction, while speakers may adjust their voice relative to their normal speaking voice, not all of them will do so in the same way. Such changes are relative to speakers' individual voices, but the key characteristic is that there is some measurable variation from the baseline acoustic features of their speech. This view was supported by a review of the literature, where previous work has found conflicting results on all measures. On this basis, it is not surprising that there was much variation in terms of how the non-serious utterances analyzed for this study varied from their serious counterparts.

However, taken in aggregate, the results of the acoustic analysis suggest several

interesting trends. A summary of the results of the acoustic analysis are presented below in Table 3.27.

Table 3.27.

Changes in Acoustic Features: Serious → Non-Serious Utterances												
	Simone Excerpts 1/2			Maximilian Excerpts 3/4			Gerhart Excerpts 5/6			Marion Excerpts 7/8		
Pitch	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg
	↓	↑	↓	↓	↑	↓	↓	↓	↓	↑	↓	↑
Pitch Variation	↓			↓			↓			↑		
Vowel Formants	/i/		/ɑ/	/i/		/ɑ/	/i/		/ɑ/	/i/		/ɑ/
	↗		↘	↙		↖	↘		↗	↙		↖
Intensity (scaled)	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg
	↑	↓	↓	↓	↑	↑	↑	↓	↓	↓	↑	↑
Intensity Variation	↑			↓			↑			↓		
Beats	↓			↑			↓			↑		
Speech Rate	↓			↓			↓			↑		

In terms of pitch, three of four speakers have decreased average pitch, decreased range of pitch, and decreased variation in pitch. Pitch values of non-serious utterances appeared to converge towards a unified pitch range, which was a result anticipated on the basis of the pilot study. In addition, the speech rate of three of four speakers decreased. The outlier for both of these acoustic features was Marion, who had increased pitch range and variation, as well as increased speech rate, for her non-serious utterances.

The analysis of vowel formants, intensity, and beats revealed a correlation with colloquial regional variations in the German language. Specifically, Simone and Gerhart, speakers of Berlin-Brandenburgish, both had lower F2 values (corresponding to backer vowels), increased

range of and variation in intensity, and decreased beats per second. Maximilian and Marion, speakers of Hessian, had the opposite results, namely, higher F2 vales (corresponding to fronter vowels), decreased range of and variation in intensity, and increased beats per second.

The results of this analysis must also be considered with reference to the results of the perceptual experiment. Given that the comparison of Marion's non-serious and serious utterances is an outlier on two acoustic measures (pitch and speech rate), we might reasonably anticipate that participants in the perceptual survey would be less likely to differentiate successfully between her serious and non-serious excerpts as compared to the other speakers. Such a result would offer strong evidence for the outlier status of her pitch and speech rate values, suggesting that decreased pitch variation and range and decreased speech rate may be especially salient acoustic features of non-seriousness. An integrated analysis of the results of both experiments is offered in Chapter 5.

The initial hypothesis that utterances treated as non-serious by participants in naturally occurring conversation would differ from surrounding utterances treated as serious appears to be supported by these data. Although statistical significance cannot be measured using this small data set, the trends in the data are suggestive that a larger study is warranted. On the basis of these results, all of the acoustic features measured for this study should be considered in future acoustic analyses of non-seriousness with a larger data set in order to determine statistical significance. Finally, the strong indications for regional variation in acoustic marking of non-seriousness should also be given special attention in any future research.

Chapter 4

Perception of Non-Seriousness

Excerpt 4.1.: Michael Feldman's *Whad'Ya Know? For March 25, 2006* (07:38 - 07:52)

001	**MF**	where you from originally
002	**TP**	uh well i- the the right way to ask that question
003		is:=
004	**MF**	=>yeah<=
005	**TP**	=where did you grow up between four and twelve,
006		(0.5)
007	**TP**	·h:: and when i was in the united [states]
		[ ]
008	**MF**	[>don't te]ll
009		me the right way to ask a question< i don't
010		tell you your job, ☺
011	**AU**	HAHAHAAHA
012	**JS**	°hehehehehe°
013	**TP**	°heh ·h:: heheh°
014	**MF**	your lingual job whatever it is °it doesn't
015		sound,° ☺

Note. Transcribed interview of Purnell and Salmons (2006) by Feldman. Audio retrieved May 7, 2012, from Michael Feldman's *Whad'Ya Know?* website, http://www.notmuch.com/Show/Archive.pl?s_id=394. Transcription is my own, including any errors or deviations from original audio.

4.1. Introduction

Although laughter can be deployed in a variety of sequential contexts to perform a variety of actions wholly apart from treating an utterance as non-serious (e.g., Jefferson, 1984), it plays a vital role in doing so: laughter is, as Sacks (1972) argues, the only response which is appropriate following a non-serious utterance. As described at length in Chapter 3, laughter in this sequential environment, also serves to obviate the need for any further serious response to whatever other action(s) the utterance may otherwise have been understood to have done (Sacks, 1972). Failure to laugh subsequent to a [+ humor] utterance may be treated as a source of trouble and subject to repair (Jefferson, 1979). Conversely, if an interlocutor laughs subsequent to a [- humor] utterance and fails to provide a serious response to the action of the utterance, this may also in turn be treated as a source of trouble and subject to repair, wherein the speaker may pursue a serious response by indicating the seriousness of their utterance (Jefferson, 2010).

Contiguity is therefore properly maintained, in the case of non-seriousness, through the deployment of laughter subsequent to the [+ humor] utterance. Given that treatment of an utterance as [+ humor] necessitates the deployment of laughter, in order to give the response properly fitted to the previous action, one must be able to first identify the [+/- humor] status of the utterance in real time. In light of the clear evidence for a preference for contiguity (Sacks, 1987), the accurate identification of the [+/- humor] status of an utterance is of tantamount importance. Particularly in the case of a [+/- humor] determination, where laughter is distinct from and takes the place of any other potential next action, as described in Chapter 3, such determination can have serious social consequences (Sacks, 1972). While the content of the utterance in a particular sequential context is *sometimes* sufficient information for a participant to correctly identify a non-serious utterance, this is not always the case. The question then remains

how participants in naturally occurring conversation are able to routinely identify [+ humor] utterances in an ambiguous sequential environment.

In parallel, ample evidence has demonstrated the central role of recipient design in naturally occurring conversation (e.g., Sacks, Schegloff, & Jefferson, 1974). Although the concept was originally hypothesized to apply chiefly to the domains of word choice, topic selection, and sequential organization, as argued in Chapter 1, it is feasible that it also extends to the phonetic domain. Indeed, some researchers have already found evidence for phonetic variation in how turns-at-talk are marked for the recipient as designed to perform a certain type of action, to date particularly in greetings (e.g., Kaimaki, 2011; Pillet-Shore, 2011; Schmitt & Knöbl, 2013) and in response tokens (Golato & Fagyal, 2008).

As such, it was hypothesized that the sensitive sequential environment of a non-serious utterance may also potentially be designed for the recipient as easily identifiable, not just sequentially but also phonetically. If this were true, then we should be able to detect such phonetic variation in the acoustic signal, as investigated in Chapter 3, and it would be the case that participants in a survey could differentiate between utterances which had been treated as non-serious versus serious in a naturally occurring conversation absent full context. In order to test whether the latter is true, as laid out already in Chapter 2, the following two hypotheses were established:

H_{1B}: Absent full conversational context, participants in a survey can identify utterances which were treated as non-serious in the original conversation at a rate greater than chance would predict.

H_{1C}: When presented with acoustic data, participants in a survey can identify utterances which were treated as non-serious in the original conversation at a greater rate of accuracy than when presented with transcripts only.

Rejection of the corresponding null hypotheses, i.e., that participants in the surveys cannot

identify non-seriousness at a rate greater than chance, would provide strong corroborating evidence in support of the argument that recipient design influences the organization of talk-in-interaction down to the phonetic level.

The perceptual experiment presented in this chapter builds on the sequential and acoustic analyses from the previous chapter by utilizing the same excerpts which were analyzed for that chapter. As Thomas (2002) demonstrates, the role of perception in sociolinguistic studies has generally been neglected and underestimated. He shows that many studies documenting sociolinguistic variation using production data ultimately fail to differentiate between production and perception. Specifically, citing the work of Kerswill and Wright (1990), Thomas argues that reliance on transcribed data, which is subject to the perceptual limitations of the transcriptionists, can result in data which actually measure analysts' perception as opposed to participants' production.

In other words, what is being measured in these studies may in fact have little to do with actual variation in *production*, but rather merely reflect the *perception* of variation on the part of the researchers themselves. Failure to consider the role that perception plays in studies intended to measure differences in production ultimately undermines the validity of the results (Thomas, 2002). Therefore, as Thomas argues, it is necessary to test not only for variation in production, but also whether said variation is perceptible and salient to other speakers of the language under study. Verification of perceptibility and salience with a perceptual experiment in turn guards against overreliance on the perceptual biases of the analyst. In that sense, the results presented in this chapter represent a necessary "next step" in the process of identifying the key acoustic properties of utterances treated as non-serious by participants in naturally occurring conversations, in line with Thomas' (2002) recommendations.

On the basis of these considerations, the data for this study were selected with the constraints of both the production and the perception experiments in mind. As described in Chapter 2, it was necessary to select excerpts where no participants laughed in overlap with the utterance which was treated as non-serious. Such excerpts can be trimmed prior to the participants' reactions in order to test whether participants in the perceptual experiment can differentiate between utterances treated as non-serious versus those treated as serious. Selecting excerpts which could be used both in the perceptual experiment presented here and in the production analysis presented in Chapter 3 allows direct comparison of the results, as well as verification of whether the features under study are perceptible and salient to the community of speakers at large, not just to the researcher(s).

After first reviewing the results of previous perceptual studies of non-seriousness (i.e., humor), I provide an overview of the data collected from the perceptual survey. I then offer an analysis of the data which examines the results under the two survey conditions and for each speakers' pair of excerpts, the reasons participants gave for their responses on the surveys, and how participants' responses vary by age, gender, and profession. Based on this analysis, I conclude that participants are able to predict at a rate greater than chance whether an utterance was treated as non-serious versus serious in a naturally occurring conversation without hearing the reactions of the participants in the original recording.

4.2. Perception and Non-Seriousness

A handful of perceptual studies have investigated whether participants are able to differentiate between serious and non-serious (or joking) utterances. Broadly speaking, these studies fall into two distinct branches of linguistics: semantics and (socio)phonetics. The long history of theories of humor in semantics and pragmatics stands in contrast to the relatively new

(socio)phonetic research (e.g., Attardo, 1994; Dynel, 2009); our focus here is on the latter. These two research traditions, although related in important ways, ultimately approach the question of how humor is done with vastly different research paradigms.

As described in Chapter 1, research on the semantics of humor draws on an array of fields, including most prominently philosophy, rhetoric, and psychology (Attardo, 1994). In these fields, the main direction of inquiry has been to describe the precise rhetorical elements which constitute a humorous text, from Plato's and Aristotle's descriptions of the poetics of comedy to Freud's analysis of humor as the release of the subconscious (as cited in Attardo, 1994). This in turn informed the semantic view of humor, which seeks to define the precise semantic properties of humorous texts as an extension of Chomsky's competence model, proposing a sort of "humor competence" as part of the underlying linguistic system (Attardo & Raskin, 1991). Influential models of this variety include the Incongruity-Resolution Model (IRM), the Isotopy-Disjunction Model (IDM), the Semantic Script Theory of Humor (SSTH), and most recently, the General Theory of Verbal Humor (GTVH). All of these have in common a focus on a disjuncture or rupture from previous context which leads to an alternate interpretation of the utterance in question, as in when a sentence such as, "That's great!" may be understood sarcastically in a context where the referent is identifiably *not* great (Attardo, 1994; Dynel, 2009). These clearly illustrate the influential role of, for example, psychology (script theory) and rhetoric (disjunction from previous context) in the development of semantic theories of humor.

The original research which defined the model for the GTVH sought to bring under one rubric all varieties of verbal humor, including, but not limited to, puns, canned jokes, word play, sarcasm, and "teasing" or "kidding" (Attardo, 1994; Attardo & Raskin, 1991). Ultimately, this

rubric harkens back to the notion that all humor can be described as a sudden disjuncture or incongruity which activates an alternate script for understanding of the utterance as it relates to the preceding context. Accordingly, all humor would in principle be perceptible solely from the content of the utterances and its surrounding context, without any need to reference prosodic information.

Studies which investigate humor following the GTVH or other similarly related research paradigms therefore generally seek to determine how some certain content interacts with a given context to create the necessary disjuncture or incongruity, such that the text in question is made identifiable as humorous or non-serious. I do not review of that full body of literature here, as it is not immediately applicable to the present study. However, given that the bulk of research on the linguistics of humor falls squarely into this category, especially as pertains to perception of humor, it is worthwhile to consider examples of such studies, and in particular how these contrast with this study.

As noted above, of particular interest in much of this work is identifying the necessary threshold of disjuncture from the preceding context which will render a sentence such as, "That's great!" as clearly incongruent so that it is understood as non-serious by participants in the study. An excellent example of such a study is Gerrig and Goldvarg (2000). The authors tested the hypothesis that situational disparity and echoic mention of preceding semantic elements both additively increase the perception of a given sentence as sarcastic (i.e., non-serious) by composing stories with a final sarcastic line. Each story was written in four different versions based with low versus high situational disparity, and with and without echoic mention. For example, in a story about writing a paper, the paper might be characterized as three pages in length (low situational disparity) versus ten pages (high situational disparity). Echoic mention

entails a previous line in the story, such as, “Your assignment is X pages, right?” followed by the sarcastic lines, “Yeah, but I can easily write it all before dessert!” (low echoic mention) versus, “Yeah, but I can easily write all X pages before dessert!” (high echoic mention). Participants then rated the final lines of the stories on a Likert scale. The authors found that both high situational disparity and high echoic mention correlated with increased perception of sarcasm, and that this effect was additive, i.e., final lines which were high on both variables were rated as more sarcastic than final lines which were high on only one of the two. The authors conclude that both variables contribute to creating the necessary incongruity or disjuncture from the preceding context which renders a given sentence as sarcastic.

Kreuz and Roberts (1995) also composed stories in a manner similar to that described above with ironic final sentences to test the perception of irony. However, their stories were designed to test two different variables, namely hyperbole and veracity. They found that highly hyperbolic sentences increased the perception of irony, as did untruthful sentences (truthfulness being determined by the content of the sentence as related to the preceding context). As in the Gerrig and Goldvarg (2000) study, these effects were additive. Finally, in a subsequent study, Kreuz and Link (2002) composed a new set of stories which also had ironic final sentences. These stories were designed to test three variables, namely positive versus negative outcome, positive versus negative evaluation of the outcome, and high versus low common ground between the characters. High versus low common ground were illustrated by indicating that the characters addressing each other in the final line shared the relationship of, for example, spouses (high common ground) versus police officer and citizen (low common ground). In this case, the authors found that negative outcome, negative evaluation, and high common ground all contributed additively to the perception of ironic intent by participants in the study.

In this overview, we can see that the goal of these and similar studies is to identify and describe the contextual variables which increase the perception of sarcasm or irony, and these studies have in fact successfully revealed much about the semantic and rhetorical features of non-seriousness. Insofar as this pertains to the present study, however, it is worth noting that all of the stories from the studies cited above were composed for the specific purpose of investigating these variables, and they therefore offer little in the way of understanding how non-seriousness is done in everyday conversation. Indeed, it is unsurprising that, if one were to compose some set of stories *intended* to be read as more or less sarcastic, the ones *intended* to be more sarcastic would then be perceived as such.

A substantially more interesting question arises when we consider that, unlike in the above experiments, we cannot know a speaker's intent when observing everyday interactions (e.g., Heritage, 1990). We must therefore consider how it is that participants in conversation are still able to navigate everyday interactions with some modicum of success, despite our inherent inability to know our interlocutor's intent without inquiring. It is easy, in other words, to intentionally compose a story with a certain context which renders a sentence non-serious, and for participants to identify it as such when reading it multiple times in an environment devoid of complicating variables. The greater mystery is how we are able to identify non-seriousness as such in real-time, despite the lack of clearly defined context and control of variables.

Furthermore, these studies have focused nearly exclusively on identifying the contextual and rhetorical features of joking, and as such, have failed to consider in their studies what role the acoustic stream plays in perception of non-seriousness. In light of this, it has been hypothesized by many researchers that manipulation of the prosodic and/or acoustic stream is used to mark a given utterance as non-serious. Indeed, Kreuz and Roberts (1995) devote nearly

half of their article to a review of the literature on the role of the "ironic tone of voice". However, they dismiss this hypothesis without testing it in their study, arguing in favor of the hypothesis that contextual information is enough to make the [+/- humor] determination, which would render any "ironic tone", in their words, "superfluous" (p. 28). Yet as noted above, it is unsurprising that they were able to construct scenarios intended to be ironic, and that these were in turn perceived as such. The experiment therefore mainly demonstrates that they were able to successfully compose ironic scenarios by manipulating their chosen contextual variables. This offers support for their argument that the selected variables contribute to creating the disjuncture required to prompt a non-serious reading of the target sentence, yet fails to support their second argument that prosodic information plays little or no role in perception of irony in everyday interactions.

Turning to (socio)phonetic research on the linguistics of humor, researchers have conducted a range of experiments intended to test whether participants are able to perceive non-seriousness using prosodic information either in conjunction with the content and surrounding context or in isolation. Such studies utilize four different types of speech signal: scripted performances of full sentences, nonsense syllables, filtered content, and interactions on talk radio shows (the last of these being of particular interest to the present study). All found at least some evidence for the ability of participants to perceive non-seriousness using prosodic information.

Among these types of studies, by far the most commonly utilized speech signal is performance of sentences from a script. For example, the stimulus materials in a study on perception of sarcasm in Cantonese and English consisted of a series of four sentences read aloud in one of four tones of voice: neutral, sincere, sarcastic, and humorous (Cheang, 2008; Cheang & Pell, 2008, 2009, 2011). Three native speakers of each language recorded the

sentences, and the authors then selected the best examples for use in the perceptual study. They found that participants were able to identify each tone of voice successfully in their native language, however, they could not do so in the unfamiliar language. While able to identify sentences in the unfamiliar language produced as neutral and sincere to some degree, participants performed at chance levels on those performed as sarcastic and humorous. The authors argue this offers evidence that phonetic cues of sarcasm and humor are salient to native speakers of the languages tested, while they are not salient to non-speakers of the languages.

Voyer and Techentin (2010) designed a similar study with one crucial difference: they used the same speaker for all recorded tokens, which offers greater control of variables in the analysis. The speaker recorded twelve sentences in three tones of voice: neutral, sincere, and sarcastic. The participants in the study were able to identify which tone of voice was intended for all sentences at a rate greater than chance. They were most successful at identifying the sincere tone, followed by sarcastic and finally, neutral. The authors conclude there is evidence that, absent any conversational context, participants can differentiate between intentional performances of neutral, sincere, and sarcastic token sentences with the same content. However, they also argue that this does not prove conclusively that there is a salient sarcastic tone of voice, as they did not include a perceptual test of written versions of the sentences in their study as a control measure.

Schaffer (1981) conducted a largely similar study, in which four speakers recorded five sentences each, first in a literal tone of voice and then in an ironic tone. Participants in the perceptual survey were able to differentiate between the literal and ironic tokens at a rate greater than chance, lending further support to the hypothesis that non-seriousness may be encoded phonetically. Schaffer's (1981) study is particularly important because, in contrast to Cheang

(2008) and Cheang and Pell (2008, 2009, 2011), she offered an analysis that divided out participants' accuracy rates by speaker, demonstrating that some speakers were more "successful" than others at performing irony. In other words, some speakers' performances were more salient to the participants in the perceptual survey, and these salient speakers had certain phonetic cues of irony in common which differed from the less successful speakers. Following her example, I offer a similar analysis in this chapter, and later present a summary of the phonetic features which the most "successful" speakers had in common in Chapter 5.

In a final example, Auberge and Cathiard (2003) employed an innovative approach in collecting their corpus of recordings and for their experimental design. They presented four speakers (three of whom were professional speakers) with a set of sentences in French to perform, each presented one at a time on a computer screen. As they progressed through the list, increasingly humorous jokes were included, with the intent of eliciting amusement. The speakers were subsequently asked to review the recordings and indicate at which points they were genuinely amused. They then performed these same sentences again with neutral prosody and facial expressions, and with neutral prosody but with a forced (or mechanical) smile.

Participants in the first perceptual survey rated the genuinely amused performances as most amused of the three, followed by the sentences produced with a mechanical smile. In a subsequent experiment designed to test the McGurk effect (McGurk & MacDonald, 1976), Auberge and Cathiard (2003) matched the audio stream of the genuinely amused tokens to the video stream of the mechanical smile tokens, and vice versa. Presented with discordant stimuli, participants in this survey tended to rate the utterances with genuinely amused video streams as more amused than those with mechanical smile video streams, but at a lower rate than when the genuinely amused video was paired with the correct audio. In other words, they tended to follow

the video stream in indicating their perception of amusement, although they rated the discordant stimuli with mechanical smile audio and genuinely amused video streams as somewhat less amused than those with correctly paired genuinely amused video and audio streams. The authors conclude that the first experiment provides further evidence for salient phonetic cues of amusement, while the second confirms that the McGurk effect holds for perception of the same.

Other sociophonetic studies on perception of non-seriousness have utilized data which did not convey any identifiable content. This may take the form of nonsense syllables or full sentences with the content filtered out. In the former category, Tartter's (1980) study is the most frequently cited example. She argued that the perception of "happy talk" or "smile voice" is directly attributable to the physiological effects of smiling (i.e., of spread, taut lips) on the vocal tract. The speakers (six total, three each male and female) in her study produced four full sentences, in addition to a series of nonsense syllables, consisting of nine variations of [hVd] nonce tokens and sixteen [Cah] nonce tokens (note that the [h] coda is not a permitted coda in English, making these truly nonsense syllables). The speakers were instructed to record the tokens first "straight-faced" (p. 25), and then with a forced smile, i.e., they were told to smile but "not to attempt to sound happy" (p. 25).

The utterances were then randomized and presented to participants in the perceptual survey under three conditions (Tartter 1980). Under the first experimental condition, participants were aware of how the tokens were produced and asked to select the one which was produced with a smile. Under the second and third conditions, participants were not made aware of how the tokens were produced. Under the former, they were instructed to select the happier token, while under the latter, they were to select the sadder token. Participants in all three survey types were able to correctly differentiate between the tokens produced with and without a smile at a

rate greater than chance, with the exception of two speakers under the third experimental condition. For the first speaker, they performed at chance levels, while with the second, they consistently misidentified the tokens, identifying the smiled tokens as non-smiled and vice versa. Tartter (1980) concludes that this study provides evidence that the effects of smiling on the vocal tract are salient to listeners, even when no genuine amusement is occurring. Furthermore, similar to Schaffer (1981), Tartter (1980) provides analysis of participants' success in identifying smile voice by speaker and how this corresponds to changes in production, demonstrating which phonetic variations are most salient (and therefore most successful) in cuing smile voice. As noted above, I return to this point in Chapter 5.

Tartter and Braun (1994) replicate and expand on these findings in a later study, wherein speakers recorded ten variations of [hVd] tokens in one of three lip positions, namely neutral, forced smile, and forced frown. They were instructed to form the lip positions without attempting to project any emotion, as in Tartter's (1980) previous study, and each speaker was recorded using all three lip positions at a normal volume and in a whisper register. The tokens were randomized to create four perceptual survey conditions: normal neutral versus smiled, whispered neutral versus smiled, normal neutral versus frowned, and whispered neutral versus frowned. In a forced choice experiment, participants selected the smiled tokens over the neutral tokens as sounding happier at a rate significantly greater than chance, and also the neutral over the frowned. Participants were slightly more successful with the tokens produced at normal volume than in whisper register. The authors conclude that this corroborates and extends the evidence from Tartter's (1980) earlier study, and lends support to the hypothesis that both smiling and frowning produce changes in speech production which are salient to untrained listeners. Tokens produced with a smile, they argue, are associated with expressions of

happiness, even when the smile is forced and not associated with any actual feelings of happiness.

Some researchers have sought to eliminate the meaning of the words as a confounding variable in their research by forcing participants to rely on the acoustic signal alone. This is accomplished by removing some phonetic features normally needed to identify phonemes, which renders the words unintelligible while preserving the bulk of the intonation curve. Bryant and Fox Tree (2005), for example, accomplish this by using high and low band-pass filtering, which attenuates frequencies above and below a given threshold. The remaining unaltered frequencies are heard as speech-like prosodic curves, but the words are unintelligible. Twelve non-ironic, twelve dry sarcastic (i.e., deadpan), and eleven "dripping" (p. 260) sarcastic utterances were first harvested from talk radio shows. The non-ironic and dry sarcastic utterances had ambiguous content, that is, the sarcastic utterances could not be identified as such without hearing them, whereas "dripping sarcasm" was defined as those utterances which had particularly exaggerated prosody and/or were unambiguously sarcastic even without hearing them. The utterances were then high and low band-pass filtered, creating two survey conditions. Participants in the surveys rated the dry sarcastic utterances as equally sarcastic to the non-ironic utterances. They performed slightly better on the high band-pass filtered set, rating the dry sarcastic utterances as slightly more sarcastic, but the difference was not statistically significant. However, they rated the instances of "dripping sarcasm" as significantly more sarcastic under both survey conditions, performing somewhat better on the low band-pass filtered set.

Rockwell (2000) similarly created survey stimuli filtered for content. The speakers (consisting of professional actors, radio announcers, and speech team competitors) were given three cards to perform, all of which contained a target utterance. The first two cards were

presented one at a time in a random order, and consisted of two different scenarios with the target utterance embedded as a quote from one of the characters towards the end. The scenarios were composed in such a way as to elicit differing performances of the target line, namely a literal performance and a spontaneously sarcastic one. The participants were allowed to study and practice each card for as long as they wished before the recording was initiated. The final card was viewed only after recording the first two scenarios. It contained only the target line with no accompanying scenario, and the instruction to record it in a sarcastic tone of voice (posed sarcasm). The content of the stimuli was then filtered at a professional sound studio; no specific technical information was provided about how the content was filtered, although the resulting signal is described as "a natural-sounding utterance, which was difficult, if not impossible, to understand" (p. 489). The recordings were then randomized to form a collection of 36 target stimuli.

In all the other perceptual studies described above, the participants completed the survey alone using a computer or in the presence of a researcher controlling an audio stream. In contrast, one unusual aspect of Rockwell's (2000) study is that she played the stimulus materials over loudspeakers to a class of over 100 students. This is noteworthy because there is the possibility that this had a negative impact on participants' ability to adequately process the signal without recourse, for example, to adjust the volume to their desired setting or replay a given token as needed. Indeed, the participants rated the tokens from the first and second cards (the literal and spontaneously sarcastic performances) as nearly equally sarcastic using a Likert scale, although the posed sarcasm from the third card was rated as significantly more sarcastic than the other two.

Taken together, the studies reviewed up to this point have all offered some evidence that

untrained listeners are able to differentiate between [+/- humor] utterances when doing so based on an acoustic signal. However, all of these studies have relied on performances of data from a script, whether full sentences or nonsense syllables, for the survey stimuli. Only Auberge and Cathiard (2003) and Rockwell (2000) attempted to elicit speech samples for use in their perceptual studies which contained at least marginally genuine expressions of amusement or sarcasm. Given that humor, irony, sarcasm, jokes, or any other forms of non-seriousness are not normally performed on command into a microphone at the behest of a researcher, it is questionable at best to what degree the salience of these performances of non-seriousness corresponds to the salience of the same in everyday talk-in-interaction. Ultimately the performances used for these sets of perceptual survey stimuli represent an idealized form of non-seriousness. Much like the stories composed for the semantic studies described earlier in the chapter, it is not altogether surprising that intentional, idealized performances of non-seriousness are salient to listeners, and successfully so.

Returning to Bryant and Fox Tree (2005), as described above, they used talk radio data for their study, which although also acted or performed for an audience, ultimately is slightly more naturalistic than data produced solely for the purpose of linguistic research, as done for the other studies. Yet the decision to use content filtering seems to have had a negative impact on participants' ability to detect sarcasm. The purpose of content filtering is to produce an impoverished signal which renders indecipherable the exact words uttered, while leaving intact *to the extent possible* some features of the acoustic signal. It is therefore not surprising that only the most obviously sarcastic utterances were salient to participants. Indeed, although the authors themselves position their results as somewhat inconclusive evidence in support of a hypothesized "ironic tone of voice", it is almost more surprising that *any* sarcasm was salient under these

impoverished conditions, given that substantial portions of the original audio sample are completely removed when using band-pass filtering.

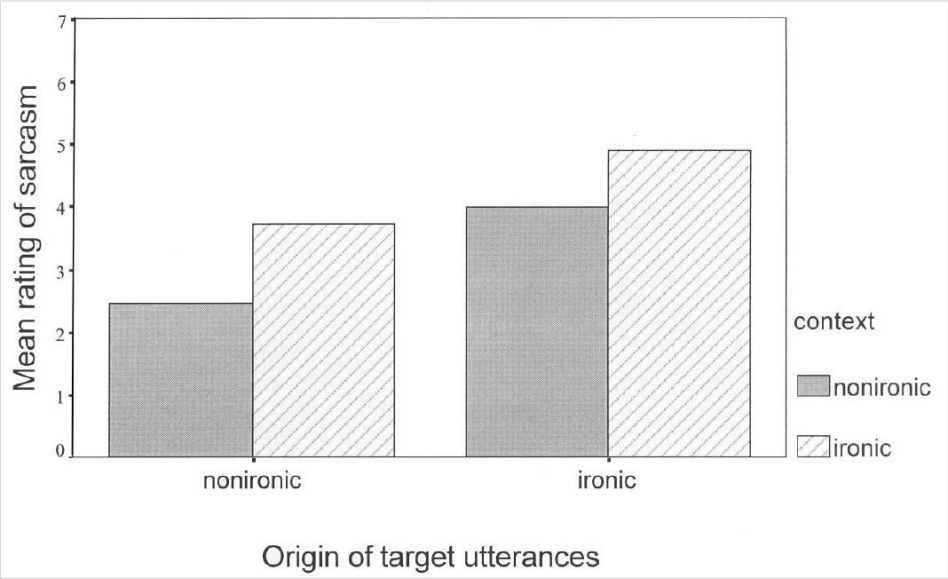
In fact, they utilized most of the same radio data in an earlier study (Bryant & Fox Tree, 2002), in which they did not use content filtering and in fact documented stronger evidence for the acoustic salience of non-seriousness as a result. In that study, they utilized only the twelve non-ironic and twelve dry sarcastic utterances. The dry sarcastic utterances are called only "ironic" in the 2002 study, but the authors note in their 2005 study that these are the same set used for the latter study, to which the "dripping sarcastic" utterances were added for the purpose of the content filtered experiment. This original set of twelve each non-ironic and ironic utterances were subjected to extensive norming procedures and then used in a particularly elegant perceptual survey. During the first norming procedure, participants presented with a words-only transcript of the stimuli rated all the utterances as nearly equally sarcastic on a Likert scale. However, participants in the second norming procedure, who listened to audio of the same utterances without a transcript, rated the originally ironic utterances as significantly more sarcastic than the non-ironic ones. Two additional norming procedures were then conducted to verify that the conversational contexts composed for use in the experiments described below were as equally believable as the original contexts.

For the first experiment, participants were presented with transcripts of two conversational contexts which both included the target utterance. One of the contexts was the original context, while the other had been composed specifically for use in the study. The latter was intended to elicit the opposite interpretation of the target utterance from its original [+/-humor] status. In other words, originally ironic utterances were provided with an additional context which prompted a literal interpretation, and vice versa, non-ironic utterances were

provided with an ironic context. Participants listened to audio of the target utterance *only*, and were asked to select which of the two contexts the utterance had most likely originated from. They correctly selected the original context at a rate significantly greater than chance would predict. In other words, by listening to the audio recording of the target utterance, they were able to differentiate between those which had originally been in an ironic context, and those which had not.

The second experiment utilized the same audio materials, but participants were only presented with one of the two contexts described above, randomized and counterbalanced to create a mixture of the original and the composed contexts. As expected from the norming procedures and the first experiment, ironic utterances in their original ironic contexts were rated as significantly more sarcastic than the non-ironic utterances in their original non-ironic contexts, using a Likert scale. The utterances in mismatched, composed contexts approached a middle ground between the utterances in their original contexts. However, the originally ironic utterances in non-ironic contexts were still rated as slightly more sarcastic than the originally non-ironic utterances in ironic contexts (see Figure 4.1., below).

Figure 4.1.



Note. "Mean responses to targets heard and read in different contexts." Reprinted from "Recognizing verbal irony in spontaneous speech," by Bryant and Fox Tree, 2002, *Metaphor and Symbol*, *17*(2), p. 109. Copyright 2002 by Lawrence Erlbaum Associates, Inc.

Out of the sociophonetic studies reviewed here, Bryant and Fox Tree (2002) offer the strongest evidence for the salience of acoustic marking of non-seriousness. The first experiment is a particularly elegant means of accessing participant's acoustic perception of a given utterance (a point to which I return in Chapter 5). Furthermore, it is striking that in the second experiment, participants still rated originally ironic utterances as more sarcastic than the originally non-ironic utterances, even when these stimuli were placed in a discordant contextual environment. If disjuncture from preceding context alone were the defining salient feature of non-seriousness, we would have anticipated that the discordant ironic context would cause the originally non-ironic utterances to be rated equally ironic as the originally ironic utterances in an ironic context, or in the reverse, that the originally ironic and non-ironic utterances would be rated as equally non-ironic in a non-ironic context. This is simply not the case. The fact that they rate the originally ironic utterances as more ironic, even when they are in a discordant non-ironic context, offers

solid evidence that there are *acoustic* features of non-seriousness which are salient to untrained listeners and enable them to make a [+/- humor] determination *despite* context.

Their study is also particularly strong because of their use of talk radio data. The vast majority of studies on the acoustics of non-seriousness, including all of those cited above save Bryant and Fox Tree (2002), have utilized stimulus materials which were composed specifically for the purpose of designing a perceptual experiment and recorded in a laboratory setting, or were artificially manipulated, as in the case of Bryant and Fox Tree (2005). These studies show that irony, sarcasm, and jokes are acoustically salient to participants when they are intentionally performed as such. However, as with the semantics studies described above, it is not altogether surprising that this should be true. Indeed, it actually would be more surprising if participants could not differentiate between sentences *intentionally performed as* literal versus sarcastic. Intentional performances ultimately draw on idealized notions of sincerity and joking, and as such, are comparatively easy to identify, as should be the case with any idealized archetype. While Bryant and Fox Tree's (2002) study represented an important step towards improving the design of stimulus materials for experiments on perception of non-seriousness, ultimately talk radio data still constitute a performance for an audience, and should not be misconstrued as fully representative of or comparable to everyday interactions.

In short, despite this extensive history of studies from the semantic and (socio)phonetic branches of linguistic research, we still know little about the perception of non-seriousness in everyday conversation. As argued above and at length in Chapter 2, this is ultimately attributable to the selection of stimulus materials. In order to test perception of non-seriousness in everyday conversation, the stimulus materials must be culled from everyday conversations, not composed and/or performed in an idealized, archetypal form. Although it is certainly clear

from the sociophonetic studies reviewed here that there is evidence in support of the hypothesis that acoustic marking of non-seriousness is salient to untrained listeners, use of stimulus materials culled from everyday conversations would offer substantially stronger evidence.

I therefore seek to fill this gap in research on perception of non-seriousness by utilizing stimulus materials culled from naturally occurring data in the manner described in Chapter 2. The results of the perceptual survey presented in the remainder of this chapter offer solid evidence in support of the hypothesis that the acoustic variation in production of non-serious utterances as compared to serious baseline utterances documented in Chapter 3 is both perceptible and salient to untrained listeners. Furthermore, these results demonstrate that such evidence can in fact be documented using stimulus materials culled from everyday talk-in-interaction, which in turn lends support in favor of the proposed methodology, a point to which I return in the next chapter.

4.3. A Note on Terminology

Throughout this chapter, I refer to participants' "accuracy rate" on their responses to the survey questions. This term requires some elaboration. First and foremost, I do not intend that survey participants' perception of the excerpts be viewed as necessarily "right" or "wrong". Just as the customer is always right, the survey participant is always right about their own perspectives on the matter at hand. In this particular context, participants' perception of what took place in each excerpt is just that: their unique perception, about which they are the authority. It cannot be assigned a value judgement in terms of correctness. In that regard, *accuracy* should not be interpreted here as referring to the correctness of their personal understanding of the excerpts, as their understanding of their own perception is inherently correct.

However, as described above, an important feature of the experimental setup is the use of naturally occurring data for the excerpts which the participants in the perceptual branch of the study viewed. The excerpts were culled from an ongoing conversation which had already occurred, i.e., the outcome of the conversation is known. Perhaps the fact of the outcome being known seems rather obvious, but it bears consideration in regard to the use of the term *accuracy*. Because there is an outcome that has already occurred and is known, it can be said that there is a correct and an incorrect answer to the question, "What will happen next in the conversation?" While it would be nearly impossible to guess at the precise words that will occur next in a given interaction with any appreciable level of accuracy, the premise of this study is that participants in a perceptual experiment are able to predict the next *type* of turn which occurred in a given interaction. For example, if a question was asked, one might reasonably predict that an answer came next, regardless of the precise form of that answer. By the same token, if one believes a joke has been made, one might reasonably predict that laughter came next.

From that perspective, there is a right and a wrong answer to the questions presented in the survey, insofar as this pertains to the recorded outcome of the conversation as it actually occurred. Although the participants, as stated above, are not correct or incorrect about how they understand and describe their own perceptions of what occurred, these same perceptions may or may not constitute an accurate description of the actual outcome of the conversations. For the purposes of this study, then, "accuracy" here refers *only* to the accuracy of their predictions with regard to the actual outcome of the recorded conversations, while emphatically *not* referring to the accuracy of their descriptions of their own individual perception of what occurred.

The use of this term is, as noted, problematic at best, which is why it is necessary to clarify its intended meaning at the outset. However, for the sake of succinctness, there is

unfortunately not a better means of expressing the fact that participants in this survey are asked to predict the potential outcome of a naturally occurring conversation which already has a known outcome, and that their prediction may or may not accurately describe the actual recorded outcome. Despite its shortcomings, *accuracy* ultimately remains the best term for characterizing these predictions. As such, for the purposes of this study, *accuracy rate* refers here to the rate at which participants were able to correctly predict the next type of turn that occurred in a naturally occurring conversation after viewing a short excerpt.

4.4. Perceptual Experiment Results

In the remainder of the chapter, I present the results of the perceptual survey, beginning with an explanation of the survey design for the two experimental groups and of the coding procedure. This is followed by an overview comparison of participants' results under both experimental conditions and over the course of the survey set. I then consider how results vary on each speakers' pair of excerpts (similar to Schaffer (1981) and Tartter (1980), as noted above), before turning to the reasons participants' gave for their responses on the surveys, and finally, how participants' responses vary by age, gender, and profession. Due to the small sample size, it was not possible to calculate statistical significance. Therefore, for purposes of this analysis, I focus on overall trends in the data as evidenced by participants' accuracy rates, and also consider the results through the lens of signal detection theory (Macmillan & Creelman, 2005). The overall trends in the data as well as signal detection results are suggestive that further study with a larger data set and statistical analysis is warranted.

4.4.1. Survey Design and Coding Procedure

In the following section describing the survey conditions and the coding procedure, I offer paraphrases of the questions and the answer options from which participants selected. The

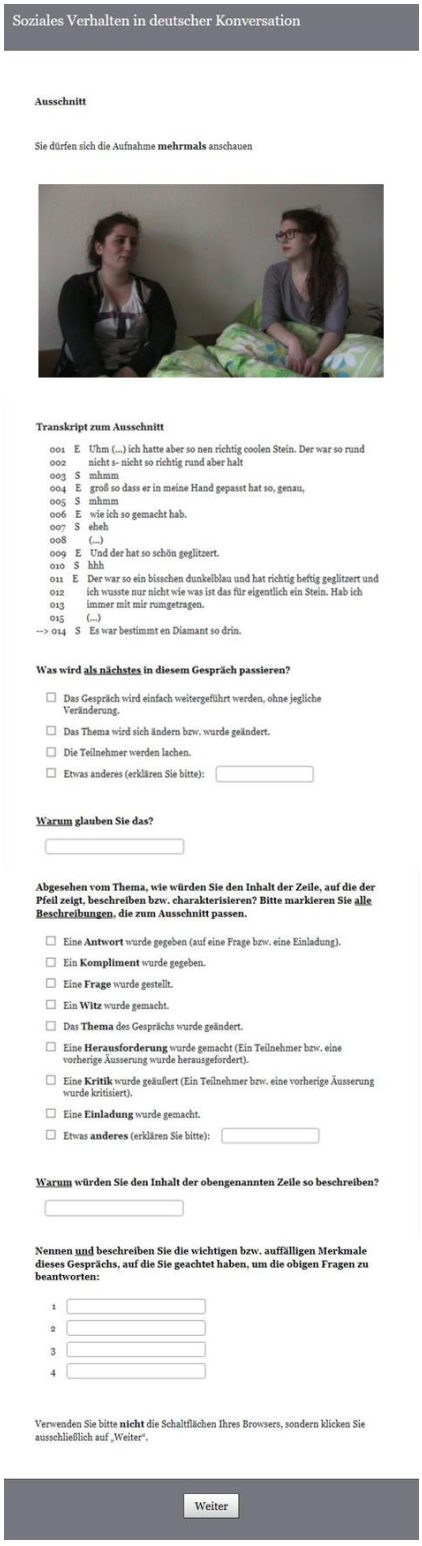
perceptual survey was conducted entirely in German, although the paraphrased versions below are given in English. The full text of the questions and the answer options as presented in the original German in the finalized live version of the survey used for this study can be found in Appendix 2.4., as well as full English translations. This appendix will help the reader understand how the analysis was conducted.

As detailed in Chapter 2, randomization was integrated into the perceptual survey at three points. First, the participants were randomly assigned to one of the two experimental groups at the equipment test page. They were not made aware of the fact that they had been assigned to experimental groups. Second, the excerpts were presented in a random order, such that the excerpts appeared in a different order for each participant and from the first survey to the second half of the experiment. Because the excerpts were not visibly numbered on the survey pages, participants were not explicitly aware that the clips were randomized, although this likely became clear to them during the second survey. Finally, the answers for the two base questions with radio button options, described below, were presented in a random order, with the "Other" option always pinned to the bottom of the list. This was done to encourage them to read the options carefully for each set of questions. Figure 4.2. below contains a screenshot of two instances of the survey questions from the second half of the survey. Note that in each image, the radio button options appear in a different, randomly determined order.

Figure 4.2.

Survey Stimuli: Randomization (Screenshots)

Survey A2, Survey B2: Example 1



Soziales Verhalten in deutscher Konversation

Ausschnitt

Sie dürfen sich die Aufnahme **mehrmals** anschauen

Transkript zum Ausschnitt

001 E Uhm (...) ich hatte aber so nen richtig coolen Stein. Der war so rund
002 nicht s- nicht so richtig rund aber halt
003 S mhmm
004 E groß so dass er in meine Hand gepasst hat so, genau,
005 S mhmm
006 E wie ich so gemacht hab.
007 S eheh
008 (...)
009 E Und der hat so schön geglitzert.
010 S hhh
011 E Der war so ein bisschen dunkelblau und hat richtig heftig geglitzert und
012 ich wusste nur nicht wie was ist das für eigentlich ein Stein. Hab ich
013 immer mit mir rumgetragen.
015 (...)
--> 014 S Es war bestimmt en Diamant so drin.

Was wird <u>als nächstes</u> in diesem Gespräch passieren?

- ☐ Das Gespräch wird einfach weitergeführt werden, ohne jegliche Veränderung.
- ☐ Das Thema wird sich ändern bzw. wurde geändert.
- ☐ Die Teilnehmer werden lachen.
- ☐ Etwas anderes (erklären Sie bitte):

<u>Warum</u> glauben Sie das?

Abgesehen vom Thema, wie würden Sie den Inhalt der Zeile, auf die der Pfeil zeigt, beschreiben bzw. charakterisieren? Bitte markieren Sie <u>alle Beschreibungen</u>, die zum Ausschnitt passen.

- ☐ Eine **Antwort** wurde gegeben (auf eine Frage bzw. eine Einladung).
- ☐ Ein **Kompliment** wurde gegeben.
- ☐ Eine **Frage** wurde gestellt.
- ☐ Ein **Witz** wurde gemacht.
- ☐ Das **Thema** des Gesprächs wurde geändert.
- ☐ Eine **Herausforderung** wurde gemacht (Ein Teilnehmer bzw. eine vorherige Äusserung wurde herausgefordert).
- ☐ Eine **Kritik** wurde geäußert (Ein Teilnehmer bzw. eine vorherige Äusserung wurde kritisiert).
- ☐ Eine **Einladung** wurde gemacht.
- ☐ Etwas **anderes** (erklären Sie bitte):

<u>Warum</u> würden Sie den Inhalt der obengenannten Zeile so beschreiben?

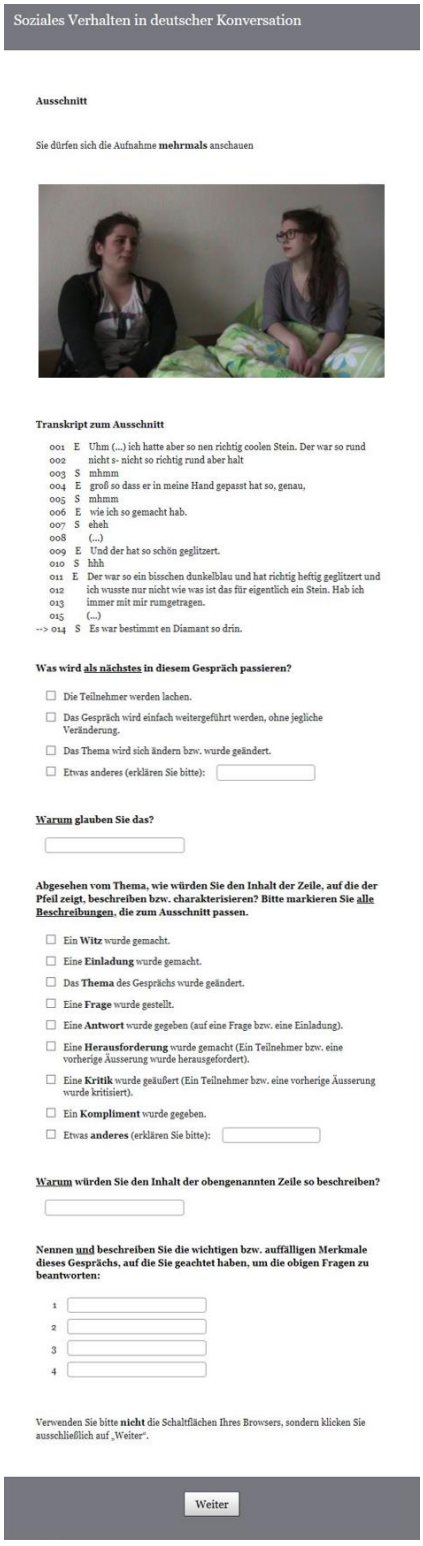
Nennen <u>und</u> beschreiben Sie die wichtigen bzw. auffälligen Merkmale dieses Gesprächs, auf die Sie geachtet haben, um die obigen Fragen zu beantworten:

1
2
3
4

Verwenden Sie bitte **nicht** die Schaltflächen Ihres Browsers, sondern klicken Sie ausschließlich auf „Weiter".

Weiter

Survey A2, Survey B2: Example 2



Soziales Verhalten in deutscher Konversation

Ausschnitt

Sie dürfen sich die Aufnahme **mehrmals** anschauen

Transkript zum Ausschnitt

001 E Uhm (...) ich hatte aber so nen richtig coolen Stein. Der war so rund
002 nicht s- nicht so richtig rund aber halt
003 S mhmm
004 E groß so dass er in meine Hand gepasst hat so, genau,
005 S mhmm
006 E wie ich so gemacht hab.
007 S eheh
008 (...)
009 E Und der hat so schön geglitzert.
010 S hhh
011 E Der war so ein bisschen dunkelblau und hat richtig heftig geglitzert und
012 ich wusste nur nicht wie was ist das für eigentlich ein Stein. Hab ich
013 immer mit mir rumgetragen.
015 (...)
--> 014 S Es war bestimmt en Diamant so drin.

Was wird <u>als nächstes</u> in diesem Gespräch passieren?

- ☐ Die Teilnehmer werden lachen.
- ☐ Das Gespräch wird einfach weitergeführt werden, ohne jegliche Veränderung.
- ☐ Das Thema wird sich ändern bzw. wurde geändert.
- ☐ Etwas anderes (erklären Sie bitte):

<u>Warum</u> glauben Sie das?

Abgesehen vom Thema, wie würden Sie den Inhalt der Zeile, auf die der Pfeil zeigt, beschreiben bzw. charakterisieren? Bitte markieren Sie <u>alle Beschreibungen</u>, die zum Ausschnitt passen.

- ☐ Ein **Witz** wurde gemacht.
- ☐ Eine **Einladung** wurde gemacht.
- ☐ Das **Thema** des Gesprächs wurde geändert.
- ☐ Eine **Frage** wurde gestellt.
- ☐ Eine **Antwort** wurde gegeben (auf eine Frage bzw. eine Einladung).
- ☐ Eine **Herausforderung** wurde gemacht (Ein Teilnehmer bzw. eine vorherige Äusserung wurde herausgefordert).
- ☐ Eine **Kritik** wurde geäußert (Ein Teilnehmer bzw. eine vorherige Äusserung wurde kritisiert).
- ☐ Ein **Kompliment** wurde gegeben.
- ☐ Etwas **anderes** (erklären Sie bitte):

<u>Warum</u> würden Sie den Inhalt der obengenannten Zeile so beschreiben?

Nennen <u>und</u> beschreiben Sie die wichtigen bzw. auffälligen Merkmale dieses Gesprächs, auf die Sie geachtet haben, um die obigen Fragen zu beantworten:

1
2
3
4

Verwenden Sie bitte **nicht** die Schaltflächen Ihres Browsers, sondern klicken Sie ausschließlich auf „Weiter".

Weiter

All participants first viewed an informed consent form, and were notified that by clicking on the "Next" button they were consenting to participate in the research. They then were given a demographics survey, followed by an equipment test page. This page contained an embedded video which participants could use to verify that their hardware and software were compatible with the survey videos, and to adjust the volume to their preferred levels (all excerpts had been scaled to have similar volume ranges in order to reduce the need for continual adjustments throughout the experiment). Once they clicked past the equipment test page, without their knowledge they were sorted into one of the two experimental groups described below, and the first survey began. After completing the second survey, participants viewed a debriefing form which revealed the true object of study, i.e., humor or non-seriousness in German. The full text of the consent form and debriefing can be found in the Appendix 2.1., whilst the demographics survey can be found in Appendix 2.5.

The conditions for the two experimental groups are summarized below in Table 4.3., and Figure 4.4. contains screenshots of the survey questions from Surveys 1 and 2 for each of the experimental groups. The screenshots are intended to provide reference for the general layout of the survey only; the full text of the small print in the screenshots can be viewed in Appendix 2.4.

Table 4.3.

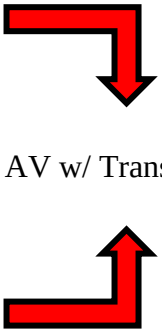
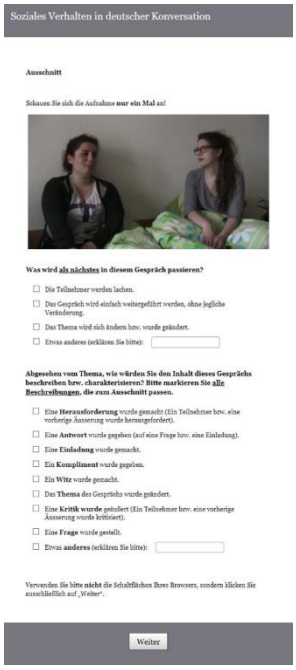
Survey Stimuli: Group A vs. Group B				
	Survey 1		**Survey 2**	
	Stimuli	**Questions**	**Stimuli**	**Questions**
Group A	AV clips only	1. What will happen next? 2. What happened in the excerpt?	AV w/ Transcripts 	1. What will happen next? 2. What happened in the excerpt? 3. Why do you believe this?
Group B	Transcripts only	1. What will happen next? 2. What happened in the excerpt? 3. Why do you believe this?		

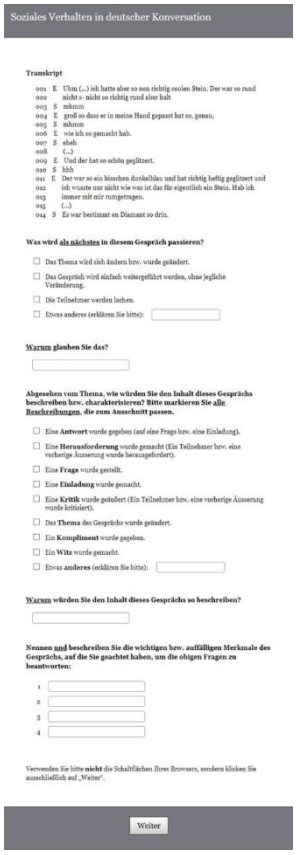
Figure 4.4.

Survey Stimuli: Group A vs. Group B (Screenshots)

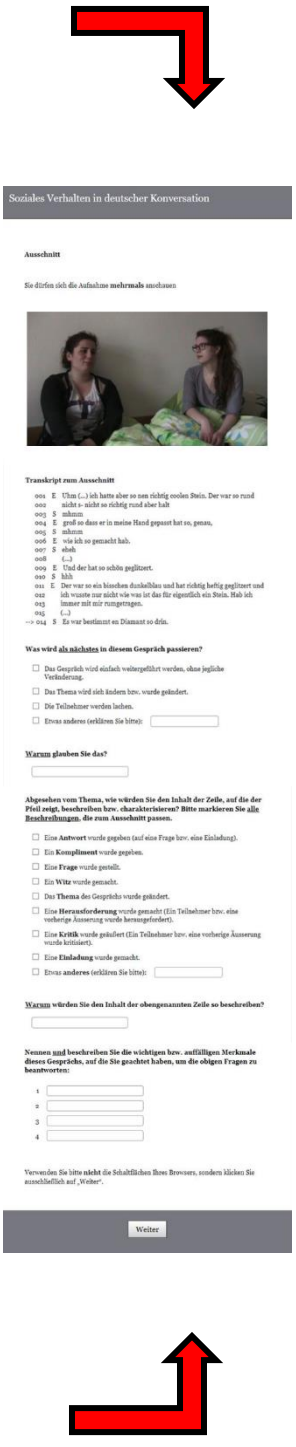
Survey A1



Survey B1



Survey A2, Survey B2



As can be seen in the table and screenshots, Group A and Group B received different surveys for the first half of the experiment, but identical surveys for the second half. Survey 1 differed in regard to the type of stimulus materials used and the number of questions. Group A viewed audio-video excerpts (Survey A1), while Group B read words-only transcripts of the same excerpts (Survey B1). Group A was instructed to view the clips only once, while Group B was not given any specific instructions about how many times they could read the transcripts, although it was assumed they would read it as many times as needed. For Surveys A2 and B2, both groups were presented with audio-video clips in addition to words-only transcripts of the excerpts, and instructed to view the videos as many times as they liked. These materials consisted of the same excerpts which both groups had viewed during Survey A1 or read during Survey B1.

The groups differed as well in terms of the questions asked about each excerpt. After viewing the audio-video clip of an excerpt, Group A was asked to answer two questions about each excerpt during Survey A1. Both of these questions were accompanied by a series of answer options from which to select, in addition to an "Other" option. As noted above, the answer options for both questions were listed in a new, random order for each excerpt in order to encourage participants to read them carefully each time. The first question asked what the survey participants predicted would happen next in the conversation. There were four total answer options: participants will laugh, topic will change, conversation will continue with no change, and "Other", this final option being accompanied by a text box for elaboration. Participants were limited by the survey to selecting only one of these four options.

The second question on Survey A1 asked what the survey participants thought had happened in the same clip, irrespective of the topic of the conversation. For this second

question, there were nine total answer options: question was asked, answer was given, invitation was given, challenge was made, compliment was given, criticism was voiced, topic was changed, joke was made, and an "Other" option, again with an accompanying text box. Participants were instructed to select as many of these options as they felt fit the interaction.

Group B was instructed to read the transcripts of the excerpts during Survey B1, and answer the same pair of base questions as Group A, i.e., what would happen next and what just happened in the excerpts. The phrasing and answer options for these questions were identical to Survey A1. In addition, participants were asked subsequent to each of the base questions why they had answered the previous question as they had, with a text box to fill as they wished. Finally, a fifth question asked them to list any interesting features of the excerpts on which they had based their answers to all of the preceding questions, with a series of four text boxes provided for this purpose. For the reasons I elaborate below, these three additional questions are summarized together as a group in Table 4.3. as the third question, "Why do you believe this?"

The questions on Surveys A2 and B2 were identical to those found on Survey B1 for both groups. What differentiates Surveys A2 and B2 from Survey B1 is the stimulus materials: both groups viewed audio-video clips and were provided with transcripts. After reviewing the excerpts and transcripts as many times as they liked, participants on Surveys A2 and B2 were asked to answer the base pair of questions about what would happen next and what had just happened in the excerpts. They were then asked the additional set of three questions, that is, why they answered the first two questions as they had and what interesting features about the interaction helped them in answering the preceding questions. To emphasize, Survey 2 was on its own identical for both Group A and Group B. The only difference was the preceding survey, that is, whether they had viewed the audio-video clips (Group A) or read the transcripts (Group

B) during Survey 1.

Despite the surface complexity of these questions and answer options, the information relevant to the present study rested solely in accessing participants' [+/- humor] determinations, i.e., whether or not a participant had indicated in response to any of the questions that a joke had occurred in a given excerpt. At its core, this perceptual experiment is a signal detection experiment (Macmillan & Creelman, 2005), wherein the participants should differentiate instances of a signal from noise, or non-instances. From the perspective of signal detection theory, the signal here refers to the non-serious utterances, while the serious utterances are the noise. The underlying question that participants are responding to is therefore, "Was a joke made in this excerpt?" A "Yes" response to this question indicates that the participant perceives that the signal is present, that is, that the excerpt is non-serious, while a "No" response indicates that they believe the excerpt is serious and therefore noise.

Given that the true goal of the perceptual survey was to determine for a given excerpt whether the survey participant believed a joke had been made, these questions and answer options may at first appear quite convoluted. However, simply asking this underlying question outright would have immediately revealed the purpose of the study, which in turn would have undermined the validity of the results. The survey questions were therefore purposefully designed to ask participants in several ways whether a joke or non-serious utterance had been made, but in such a way as not to reveal the purpose of the study. This was accomplished by presenting a variety of options for answers, such that the options for participant laughter and making a joke are buried in a set of distractors.

The coding procedure therefore ignored which distractors participants may have selected beyond whether they had indicated that a joke had been made. In practical terms, this meant that

all responses to each of the questions above could be reduced to a "Yes" or "No" response to the underlying question of whether a joke had been made. For example, for the first question about what would happen next in the excerpt, if a survey participant checked that the participants in the conversation would laugh, then this was coded as a "Yes" response, i.e., a joke was made. If this box was not checked, this was coded as a "No" response (pending analysis of any text entered if the "Other" option had been selected, as described below).

Similarly, for the second question about what had happened in the excerpt, if a survey participant checked that a joke had been made, this was also coded as a "Yes" response. Although the participants were free to select as many of the answer options on the second question as they wished, regardless of how many other boxes had been checked or not, if the joke box had been checked, then this was a "Yes" response. If the joke box had not been checked, then this was a "No" response.

In addition, if the "Other" option was checked for either of these two questions, the text box entries were analyzed to determine if the survey participant had indicated in their response that a joke was made. As an example, Participant 316 provided the following set of responses to Excerpt 7 (a non-serious excerpt) on Survey A1:

Table 4.5.

Response to Excerpt 7 on Survey A1: Participant 316	
Original Responses	**Translation**
Was wird als nächstes in diesem Gespräch passieren? Die Teilnehmer werden lachen. Abgesehen vom Thema, wie würden Sie den Inhalt dieses Gesprächs beschreiben bzw. charakterisieren? Bitte markieren Sie alle Beschreibungen, die zum Ausschnitt passen. Eine **Herausforderung** wurde gemacht (Ein Teilnehmer bzw. eine vorherige Äusserung wurde herausgefordert). Eine **Kritik wurde** geäußert (Ein Teilnehmer bzw. eine vorherige Äusserung wurde kritisiert). Etwas **anderes** (erklären Sie bitte):: Ironie, wenn man das als Witz verstehen will, ist es ein Witz	What will happen next in this conversation? The participants will laugh. Aside from the topic, how would you describe and/or characterize the contents of the conversation? Please mark all the responses that match the excerpt. A challenge was made (a participant and/or a previous statement was challenged) A criticism was voiced (a participant and/or a previous statement was criticized) Something else (please explain): *Irony, if one wants to consider that a joke, then it's a joke*

In response to the first question about what would happen next, Participant 316 indicated that the participants would laugh, which was coded as a "Yes" response. However, for the second question about what had happened in the excerpt, they did not select that a joke was made. Instead, they opted for challenge and criticism, in addition to the "Other" selection. In their clarification for "Other", they wrote: "Irony, if one wants to consider that a joke, then it's a joke (*Ironie, wenn man das als Witz verstehen will, ist es ein Witz*)". This is an excellent example of an instance where the participant did not select the joke answer from the list on the second

question, yet still entered some response in the "Other" text box which could be clearly understood as indicating that a joke was made. Any responses of this sort which stated that an utterance was ironic, sarcastic, funny, joking, or some other variant on these was therefore coded as a "Yes" response, even if they did not select the joke answer from the list. If a participant's response in the "Other" text box did not unambiguously state something of this nature, then it was coded as a "No" response, i.e., a joke was not made (unless of course for the second question they had also checked the joke response, which superseded any text entered for the "Other" option).

As noted above, the two questions about their reasons for selecting their answers to the two base questions and the final question about the interesting features of the excerpt they had noticed in answering all of the questions were bundled into one response for the purposes of the analysis. This was done because all of these questions were a package designed to access their response to one underlying question, that is, why they responded as they did. Most participants entered very similar responses to all three of the questions, or opted to enter a response to only one of them which addressed the reasons why they responded as they did, often with some text entered in the remaining boxes indicating to the reader to "see above response" or similar.

These responses were then coded as a bundled response in two ways. First, the text was analyzed to determine if they had indicated in any of the three responses that a joke had been made, much in the same manner as for the "Other" option for the first two base questions above. If they stated that it was ironic, sarcastic, and so on, then it was coded as a "Yes" response; if they did not make any statements to that effect, it was coded as a "No". In addition to coding for their response to this underlying question about whether or not a joke had been made, their responses were tagged by response category, such as body language or sound, depending on the

types of reasons given in their responses. The specific categories used and how they were determined is addressed later in the chapter, in Section 4.4.5.

Each of the responses to the questions were coded this way for all participants, resulting in a total of two "Yes" or "No" responses per excerpt on Survey A1, and a total of three "Yes" or "No" responses per excerpt on Surveys A2, B1, and B2. Full completion of the survey for Group A therefore consists of a total of 40 responses (eight excerpts on two surveys with two responses per excerpt on the first survey and three responses per excerpt on the second), and for Group B a total of 48 responses (eight excerpts on two surveys with three responses per excerpt on both surveys).

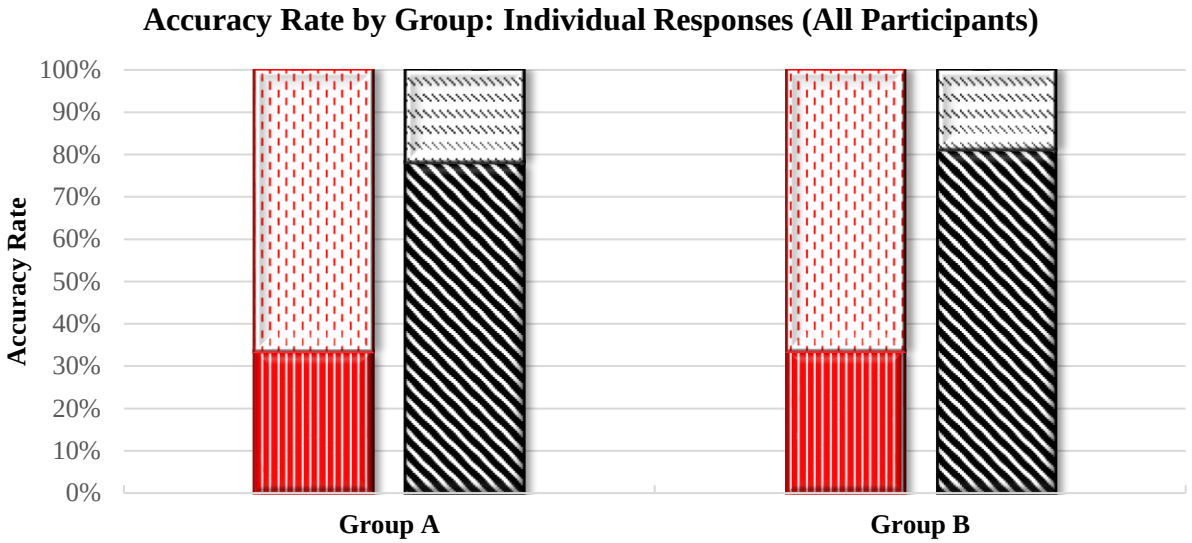
Returning briefly to the analysis from Chapter 3, recall that there were eight total excerpts used in the survey. There were two excerpts from each of four speakers, of which one was non-serious and one was serious, as determined using sequential analysis. The non-serious excerpts were assigned odd numbers, while the serious ones were assigned even numbers. A response coded as "Yes, a joke was made" to any of the three questions described above is therefore *accurate* on non-serious Excerpts 1, 3, 5, and 7, while the same "Yes" response is *inaccurate* on serious Excerpts 2, 4, 6, and 8. Conversely, a response coded as "No, a joke was not made" to any of the three questions described above is *accurate* on serious Excerpts 2, 4, 6, and 8, while the same "No" response is *inaccurate* on non-serious Excerpts 1, 3, 5, and 7. The coded responses could then be taken over the total number of responses to arrive at an accuracy rate (which, as explained above, refers only to whether or not the perceptual survey participants accurately identified that a joke had been made according to what actually occurred in the original recording, not to the accuracy of their own perception as such) for the purposes of the analysis, the results of which are presented in the remainder of the chapter.

4.4.2. Experimental Groups A and B

Turning to the results, we first examine the overall accuracy rates for both experimental groups on the two surveys in aggregate. According to the coding procedure described above, all participants' responses to the questions for each excerpt were initially considered to be individual responses independent of each other. For example, if a given participant on Survey A1 responded that Excerpt 1 was a joke for the first question, but then did not indicate that it was a joke for the second question, then this was coded as one "Yes" response and one "No" response. The former response would be accurate for Excerpt 1 (a non-serious excerpt), whereas the latter one would be inaccurate. In short, the two responses had no bearing on each other, on this view of the data.

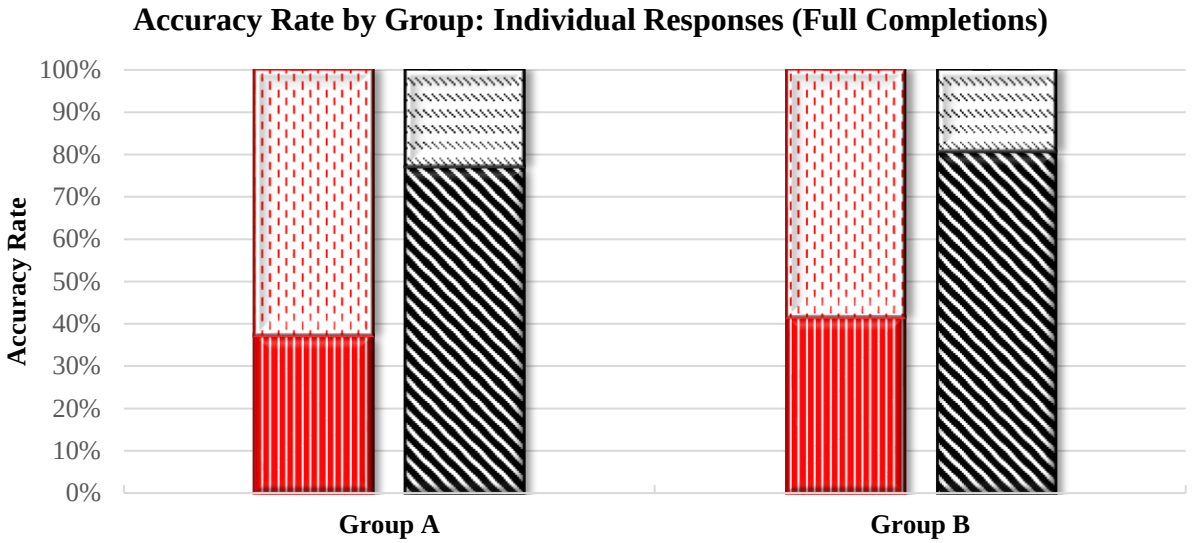
The results of this initial analysis, tallied as individual responses, are presented below in Figures 4.6. and 4.7. A brief explanation of how to read the charts and tables throughout the chapter is provided below this first set of figures.

Figure 4.6.



Group	Total N	Excerpt	PA: Is Joke?		Legend		% Correct
A	1042	Non-Serious	Yes	350	Hit		33.6%
			No	692	Miss		
	1044	Serious	Yes	227	False Alarm		78.3%
			No	817	Correct Rejection		
B	567	Non-Serious	Yes	191	Hit		33.7%
			No	376	Miss		
	574	Serious	Yes	108	False Alarm		81.2%
			No	466	Correct Rejection		

Figure 4.7.



Group	Total N	Excerpt	PA: Is Joke?		Legend		% Correct
A	670	Non-Serious	Yes	250	Hit		37.3%
			No	420	Miss		
	671	Serious	Yes	153	False Alarm		77.2%
			No	518	Correct Rejection		
B	303	Non-Serious	Yes	126	Hit		41.6%
			No	177	Miss		
	302	Serious	Yes	58	False Alarm		80.8%
			No	244	Correct Rejection		

Figure 4.6. contains the results for all participants who responded to at least one question on the survey subsequent to the demographic survey. Figure 4.7. contains the results for full completions only, that is, the participants who completed in their entirety both surveys in the set for the respective experimental group. As noted above, for Group A, this amounts to 40 total responses per participant on both surveys, while for Group B, full completion entails 48 total responses. The table legends below each chart, and throughout the rest of the chapter, indicate which survey group the participants were assigned (Group A or B), followed by the total number

of responses (Total N), the type of excerpt (non-serious or serious), and their response to the question of whether a joke had been made ("Yes" or "No"). This is followed by the total number of participant responses indicating "Yes" versus "No". Correct responses are in the cells colored green, while incorrect responses are in the cells colored red. The legend indicates which bars in the accompanying charts correspond to the accuracy rates listed in the final column of the table; the bottom half of the columns in the charts indicate the percentage of correct responses, with red for non-serious and black for serious excerpts. The final column contains the percentage of correct responses (i.e., hits and correct rejections), calculated by dividing the number of correct responses in the corresponding green squares over the total number of responses.

In terms of accuracy, both sets of participants (i.e., all participants and the subset of participants who completed the entire survey) performed better at identifying serious utterances as compared to non-serious utterances. That is, participants generally did not indicate that a joke had been made when the utterance was not treated as non-serious in the original conversation, yet they also generally did not indicate the reverse, that a joke had been made when the utterance had in fact been treated as non-serious. This can be seen in the rather low accuracy rates in Figures 4.6. and 4.7. for non-serious utterances across both participant sets on both surveys, while accuracy rates were much higher for serious utterances. This overall trend remains the same for both participant sets, although the participants who completed the entire survey performed slightly better on the non-serious utterances.

In line with the analytical framework of signal detection theory (Macmillan & Creelman, 2005), the terms *hit*, *miss*, *false alarm*, and *correct rejection* in the above table legends and throughout the chapter refer to how participants responded in relation to whether or not the signal was actually present. That is, for each excerpt, the signal was either present (non-serious

utterance) or absent (serious). Following the coding procedure described above, participants then indicated that the signal was either present or absent, which produces four possible permutations of responses, as illustrated in Table 4.8., below.

Table 4.8.

Signal Detection: Response Type Permutations		
	Respond Signal Present (Participant Answer: Yes)	**Respond Signal Absent (Participant Answer: No)**
Stimulus Signal Present (Signal: Non-Serious)	Hit	Miss
Stimulus Signal Absent (Noise: Serious)	False Alarm	Correct Rejection

Based on this framework, we can evaluate participants' ability to distinguish between utterances treated as non-serious (signal) and serious (noise) by measuring how sensitive participants were to the presence or absence of the signal. This is calculated using the hit and false alarm rates to determine *A′*, which is a nonparametric index of sensitivity ranging from a value of 0 to 1. A higher value for *A′* indicates greater sensitivity, meaning that participants were good at differentiating between signal and noise, while a lower value indicates that they are less able to detect a difference between the two sets of stimuli. A value at or near 0.5 indicates no signal detection, or an inability to discriminate between signal and noise, such that participants may have performed equally well without listening to the stimuli and simply selecting responses at random. Results less than 0.5 are an indication of signal confusion, as in mistaking noise for signal presence or vice versa. It should be noted here that *A′* is not a measure of statistical significance. In addition, *A′* is an index of sensitivity, not a unit of measurement. In other words, *A′* allows comparison of sensitivity to the stimulus *relative to* other participants in the

same experiment, but cannot be used to compare the results to other experiments.

By this measure, all participants displayed some sensitivity to the non-serious utterances, despite the rather low accuracy rates. From Figure 4.6., we find that the A group with video clips on the first half of the survey performed slightly worse ($A' = 0.626$) than the B group with transcripts ($A' = 0.656$). The participants who completed the full survey, in Figure 4.7., were slightly more sensitive to non-seriousness, while the A group continued to perform slightly worse ($A' = 0.644$) than the B group ($A' = 0.704$). These results were somewhat surprising, as it was anticipated that participants would be more sensitive to non-seriousness overall, and also that the A group would perform better than the B group. However, it is not surprising that participants who completed the entire survey would perform better. There could be multiple reasons for this, such as: participants may have improved their performance over the course of the survey; participants who completed the survey were potentially more invested in listening to the clips carefully and offering considered responses to the survey items; or participants who found the task easier than others may have been more inclined to continue the survey to full completion.

This does not take into account, however, the effects of response bias, that is, whether there was a tendency to respond preferentially with “Yes” or “No”, regardless of the actual presence or absence of the signal. This can be determined using B'', which is a nonparametric measure of the degree of response bias. The value of B'' ranges from -1 to 1, with the middle point at 0 indicating no response bias at all. An ideal participant does not have any response bias, as this would mean they are able to distinguish between signal and noise without the influence of any other factors, such as mental state or a tendency to over- or underestimate the presence of the signal. However, this ideal is largely unattainable under real experimental

conditions, so we simply want *B″* to have a value as close as possible to 0; in conjunction with a large value for *A′*, this would indicate high sensitivity to the signal (in this case, non-seriousness) with minimal influence on responses aside from the signal proper. A positive value for *B″* indicates a bias for responding with "No", i.e., that the signal is absent, while a negative value indicates a bias for responding with "Yes".

In terms of response bias, we find that both the A and B group participants in Figure 4.6. had a tendency to respond "No", that is, that an utterance was serious (*B″* = 0.135 and 0.188, respectively). This holds for the full completion participants in both groups from Figure 4.7. as well (*B″* = 0.141 and 0.220). This result indicates that, while participants were sensitive to the differences between the non-serious and serious utterance as argued above, they were also more likely to indicate that an utterance was serious overall, regardless of the presence or absence of the signal.

According this view of the data, where participant responses to the core three questions were taken as individual responses independent of each other, we find that participants were indeed able to differentiate between utterances treated as non-serious versus serious. However, their accuracy in doing so was quite low, which in combination with the tendency to indicate that utterances were serious overall leads to the conclusion that they were not especially sensitive to differences between the non-serious and serious utterances. From this perspective, it seems on initial examination that there was not strong evidence in support of the hypothesis that participants could differentiate between non-serious and serious utterances.

However, after further analysis, it became clear that viewing participant responses independently of the complete bundle of responses did not offer an accurate representation of their interpretations of the interactions. As described above, participants responded to three

questions for each excerpt. Yet in the course of coding their responses, I encountered numerous instances where participants indicated that a joke had been made in the excerpt in response to only one or two of the questions, but not all three. This has the effect of artificially reducing their rate of accuracy on non-serious utterances. If a participant has already indicated on the survey page that a joke was made in response to at least one question, it cannot be argued that they did not interpret the excerpt as non-serious. Furthermore, it may be the case that participants did not indicate on multiple questions below a single excerpt that a joke had been made for reasons wholly outside of their interpretation of the utterances, such as a desire to avoid duplicate responses or to provide more information about their interpretations aside from non-seriousness.

Indeed, there were many instances where participants indicated at least once in their response that a joke had been made, yet they did not do so in response to all of the questions for that particular excerpt. For example, Participant 184 offers the following in response to non-serious Excerpt 3 on Survey A2:

Table 4.9.

Response to Excerpt 3 on Survey A2: Participant 184	
Original Responses	**Translation**
Was wird als nächstes in diesem Gespräch passieren? Das Gespräch wird einfach weitergeführt werden, ohne jegliche Veränderung. Warum glauben Sie das? E hat gerade gelacht, aber nicht allzu herzhaft, wird also vermutlich nicht sofort weiterlachen; M wird den ironischen Ton seiner Aussage nicht durch Lachen brechen Add Tag to Answer Abgesehen vom Thema, wie würden Sie den Inhalt der Zeile, auf die der Pfeil zeigt, beschreiben bzw. charakterisieren? Bitte markieren Sie alle Beschreibungen, die zum Ausschnitt passen. Ein **Witz** wurde gemacht. Warum würden Sie den Inhalt der obengenannten Zeile so beschreiben? eindeutig ironische Beschreibung des Spielstils Add Tag to Answer	What will happen next in the conversation? The conversation will simply continue, without any change. Why do you believe this? *E just laughed, but not very heartily, so he'll presumably not continue laughing right away; M will not break the ironic tone of his statement by laughing.* Aside from the topic, how would you describe and/or characterize the contents of the of the lines which are indicated by the arrow? Please mark all the responses that match the excerpt. A joke was made Why would you describe the contents of the lines indicated above in this way? *Unequivocally ironic description of the playing style.*

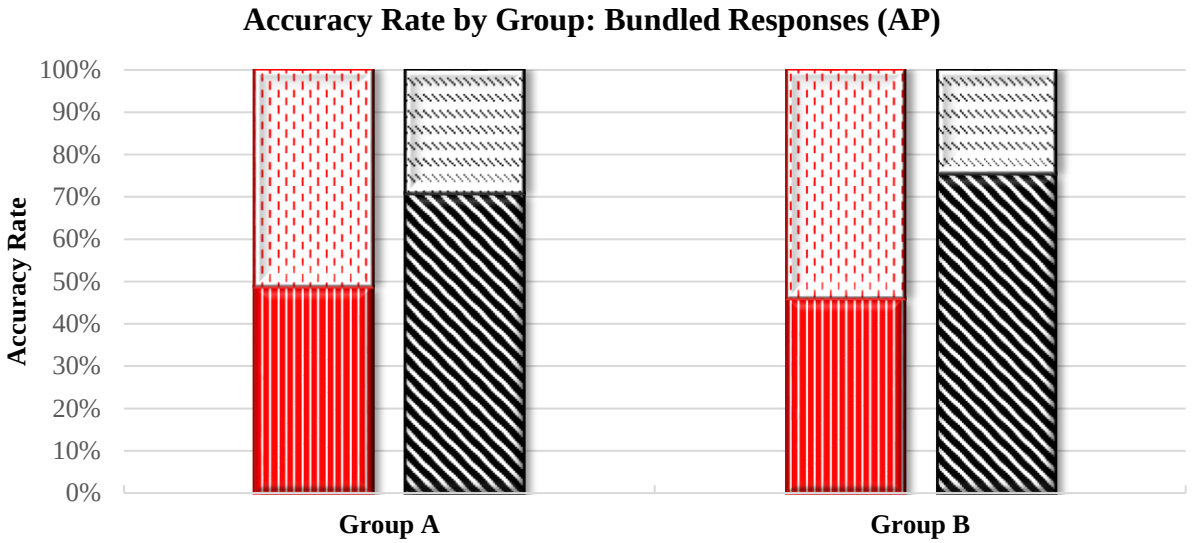
In this instance, the participant indicates that a joke was made on the second multiple choice question, but did not indicate that laughter was next on the first one. In response to the questions

about why they answered as they did, this participant first states that "M will not break the ironic tone of his statement with laughter" ("*M wird den ironischen Ton seiner Aussage nicht durch Lachen brechen*"), and then states that it is an "unambiguously ironic description of the play style" ("*eindeutig ironische Beschreibung des Spielstils*"). In this case, following the original analytical scheme, this participant's responses would be coded and analyzed as two "Yes" responses for the second and third questions, meaning that a joke had been made, and one "No" response for the first question. This had the effect of artificially lowering this participant's accuracy rate, in that it would be recorded not as one correct response, but rather as two correct responses and one incorrect. This is not reflective of their actual performance, however, especially given that it was made amply clear that the participant had indeed successfully identified Excerpt 3 as an utterance which had been treated as non-serious.

Numerous similar instances lead to the conclusion that participants' survey responses should not be viewed as discrete items, but rather should properly be analyzed as a bundle of responses. According to this view of the data, all responses pertaining to a given excerpt were analyzed as a single bundled response to three sub-questions. If a participant responded to any one of the three questions on a given excerpt that a joke had been made, then this was recorded as a "Yes, a joke was made" response for that excerpt. This was done to better reflect the true nature of responses similar to the set in Figure 4.9., wherein it is made abundantly clear that the participant has accurately identified non-seriousness, but did not indicate this on all three of the questions.

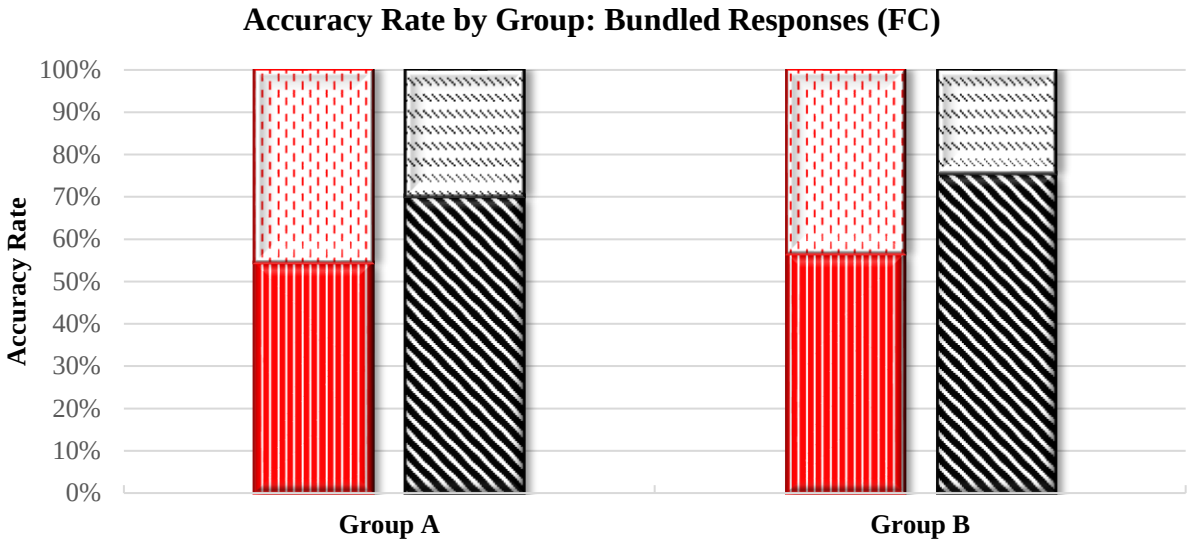
As such, we now turn to an analysis of the data as a set of bundled responses. The overall results for both survey groups are presented in Figure 4.10. and 4.11., below.

Figure 4.10.



Group	Total N	Excerpt	PA: Is Joke?		Legend		% Correct
A	522	Non-Serious	Yes	255	Hit		48.9%
			No	267	Miss		
	527	Serious	Yes	153	False Alarm		71.0%
			No	374	Correct Rejection		
B	286	Non-Serious	Yes	132	Hit		46.2%
			No	154	Miss		
	290	Serious	Yes	71	False Alarm		75.5%
			No	219	Correct Rejection		

Figure 4.11.



Group	Total N	Excerpt	PA: Is Joke?		Legend		% Correct
A	336	Non-Serious	Yes	183	Hit		54.5%
			No	153	Miss		
	336	Serious	Yes	100	False Alarm		70.2%
			No	236	Correct Rejection		
B	152	Non-Serious	Yes	86	Hit		56.6%
			No	66	Miss		
	152	Serious	Yes	37	False Alarm		75.7%
			No	115	Correct Rejection		

When viewed as bundled responses, all participants (AP) have markedly improved accuracy rates under both survey conditions on the non-serious excerpts, as can be seen in Figure 4.10., with accuracy rates over 45%. This result holds for the subset of participants who completed the entire survey (FC), as seen in Figure 4.11., where participants achieved accuracy rates over 50% on non-serious excerpts. There is some corresponding reduction in accuracy on serious excerpts, as there were certainly instances where participants had indicated on at least one of the questions for a serious excerpt that a joke had been, yet accuracy rates remain over 70% on serious

excerpts.

As under the previous view of the data, Group A displays less sensitivity to non-seriousness (A' = 0.671) than Group B (A' = 0.689) overall. The full completion group again performs slightly better by this measure, with Group A faring worse (A' = 0.701) than Group B (A' = 0.749). However, these scores are across the board better than under the previous view of the data as discrete responses. This is a clear result of taking into consideration the bundle of responses which in turn corrects for the artificial reduction in their accuracy rates on non-serious utterances which was the result of the original analytical scheme. Participants again tended to preferentially respond with "No", as under the previous view of the data, especially in Group B (B'' = 0.096 (A:AP), 0.147 (B:AP), 0.085 (A:FC), and 0.143 (B:FC)), although to a lesser degree as under the previous view of the data.

Based on this initial analysis, I first conclude that overall, Group B seems to outperform Group A, a somewhat unexpected result which warrants further analysis throughout the remainder of the chapter. In addition, we found that participant responses are best viewed as a set of bundled responses to a single excerpt rather than as discrete responses to individual questions. While this produces some artificial reduction in accuracy rates on serious excerpts, analyzing the coded data as bundled responses clearly offers a more accurate reflection of participants' reported perception of non-seriousness. I therefore abandon the original analytical scheme in favor of the bundled response view. As such, the remainder of the chapter focuses solely on analysis of the data as bundled responses.

4.4.3. Surveys 1 and 2

As noted in the previous section, Group B tended to be more sensitive to the differences between non-serious and serious utterances than Group A. We now consider how each group

performed on Survey 1 versus Survey 2. Recall that Survey 2 is essentially identical for both groups, in that they were both presented with video clips and transcripts of the excerpts. The sole difference is in the preceding survey condition: Group A viewed video clips of the excerpts without transcripts (which they were instructed to view only once), while Group B was presented only with transcripts of the excerpts (which they were permitted to read as many times as they wished).

The results for all participants (AP) on both surveys are presented in Figure 4.12., while the results for full completions only (FC) are presented in Figure 4.13., below.

Figure 4.12.

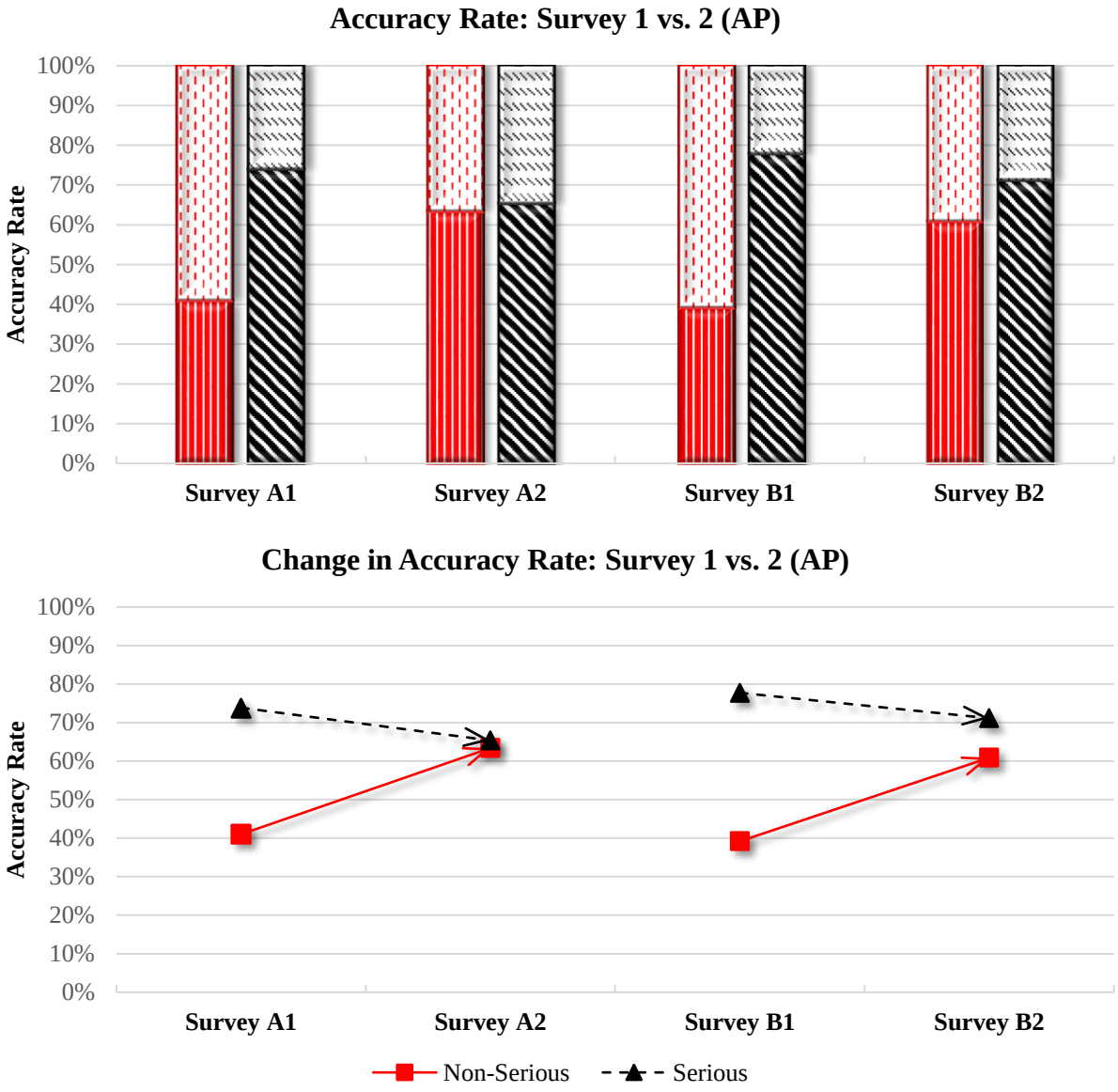

Accuracy Rate: Survey 1 vs. 2 (AP)
Accuracy Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Survey A1
Survey A2
Survey B1
Survey B2
Change in Accuracy Rate: Survey 1 vs. 2 (AP)
Accuracy Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Survey A1
Survey A2
Survey B1
Survey B2
Non-Serious
Serious

Figure 4.12. (cont.)

Group	Total N	Excerpt	PA: Is Joke?		Legend		% Correct
A1	339	Non-Serious	Yes	139	Hit		41.0%
			No	200	Miss		
	348	Serious	Yes	91	False Alarm		73.9%
			No	257	Correct Rejection		
A2	183	Non-Serious	Yes	116	Hit		63.4%
			No	67	Miss		
	179	Serious	Yes	62	False Alarm		65.4%
			No	117	Correct Rejection		
B1	194	Non-Serious	Yes	76	Hit		39.2%
			No	118	Miss		
	193	Serious	Yes	43	False Alarm		77.7%
			No	150	Correct Rejection		
B2	92	Non-Serious	Yes	56	Hit		60.9%
			No	36	Miss		
	97	Serious	Yes	28	False Alarm		71.1%
			No	69	Correct Rejection		

Figure 4.13.

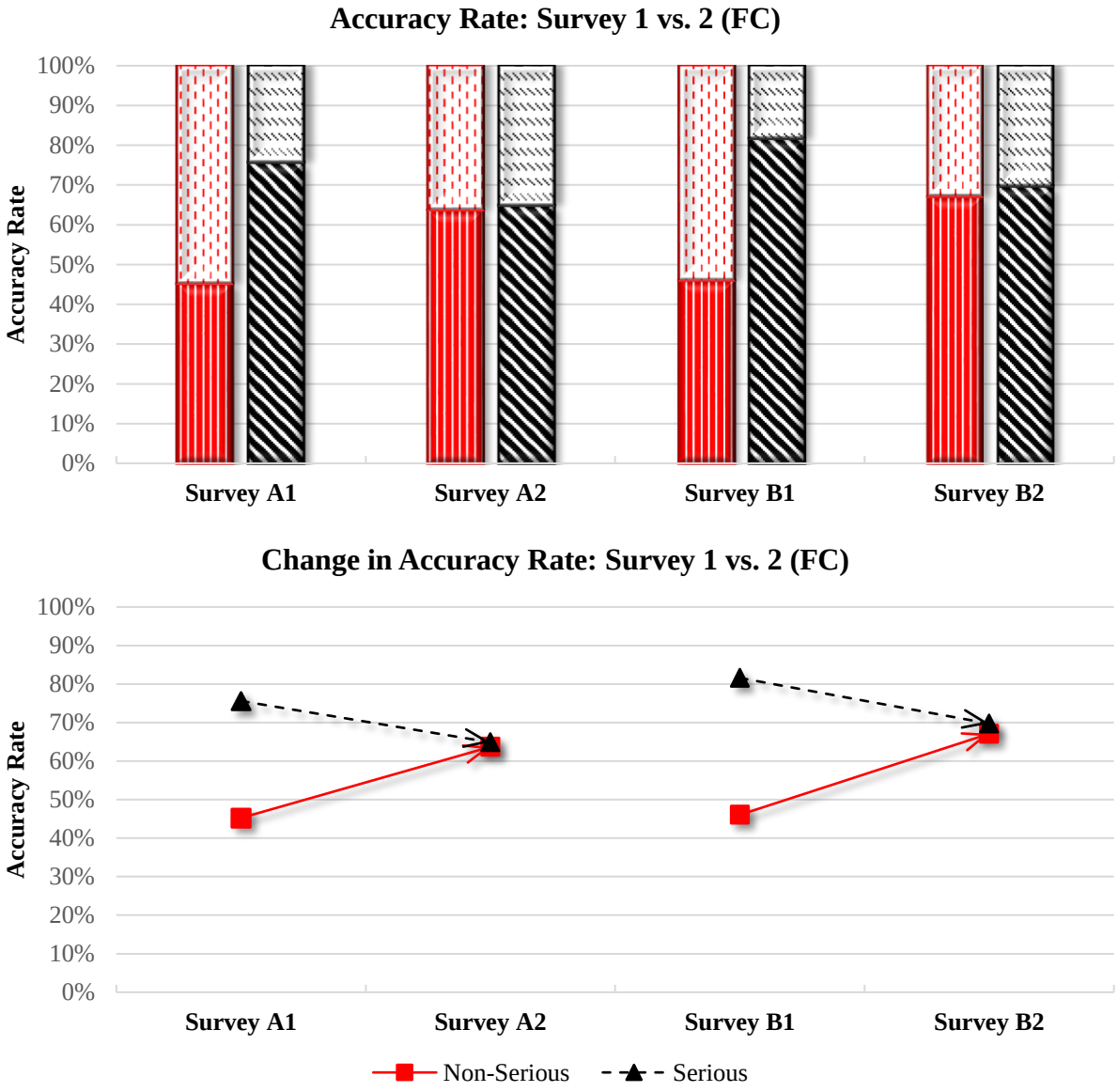

Accuracy Rate: Survey 1 vs. 2 (FC)
Accuracy Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Survey A1
Survey A2
Survey B1
Survey B2
Change in Accuracy Rate: Survey 1 vs. 2 (FC)
Accuracy Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Survey A1
Survey A2
Survey B1
Survey B2
Non-Serious
Serious

Figure 4.13. (cont.)

Group	Total N	Excerpt	PA: Is Joke?		Legend		% Correct
A1	168	Non-Serious	Yes	76	Hit		45.2%
			No	92	Miss		
	168	Serious	Yes	41	False Alarm		75.6%
			No	127	Correct Rejection		
A2	168	Non-Serious	Yes	107	Hit		63.7%
			No	61	Miss		
	168	Serious	Yes	59	False Alarm		64.9%
			No	109	Correct Rejection		
B1	76	Non-Serious	Yes	35	Hit		46.1%
			No	41	Miss		
	76	Serious	Yes	14	False Alarm		81.6%
			No	62	Correct Rejection		
B2	76	Non-Serious	Yes	51	Hit		67.1%
			No	25	Miss		
	76	Serious	Yes	23	False Alarm		69.7%
			No	53	Correct Rejection		

Comparing the accuracy rates from Figure 4.12. to those from Figure 4.13., we find that the participants in both Group A and B who completed the entire survey set (FC) generally had higher accuracy rates on Survey 1 for both non-serious and serious excerpts as compared to all participants (AP). Correspondingly, the full completion participants display greater sensitivity to non-seriousness on Survey 1 (A' = 0.684 (A1:FC) and 0.735 (B1:FC)) than all participants (A' = 0.641 (A1:AP) and 0.662 (B1:AP)). Regardless of completion status, all participants tended to have a response bias towards "No" on Survey 1, with little variation in that regard (B'' = 0.112

(A1:AP), 0.158 (B1:AP), 0.146 (A1:FC), and 0.246 (B1:FC)).

The full completion participants in Group A performed similarly to all participants on Survey 2, with accuracy rates nearly identical for both non-serious and serious excerpts. Completion status for Group A on Survey 2 had little effect on sensitivity to non-seriousness (A' = 0.723 (A2:AP) and 0.722 (A2:FC)) or on response bias (B'' = 0.012 (A2:AP) and 0.007 (A2:FC)). On the other hand, the Group B full completion participants had much higher accuracy on non-serious excerpts on Survey 2 as compared to all participants. This translates to increased sensitivity to non-seriousness for full completion participants on Survey B2 (A' = 0.769) compared to all participants (A' = 0.744), and slightly decreased response bias towards "No" among full completion participants (B'' = 0.022) compared to all participants (B'' = 0.074).

The differences between how all participants and full completion participants performed on the second survey in Group A versus Group B are likely attributable to differences in patterns and rates of attrition over the course of the survey. Figures 4.14. and 4.15., below, contain the total number of participants who completed a given number of excerpts by experimental group.

Figure 4.14.

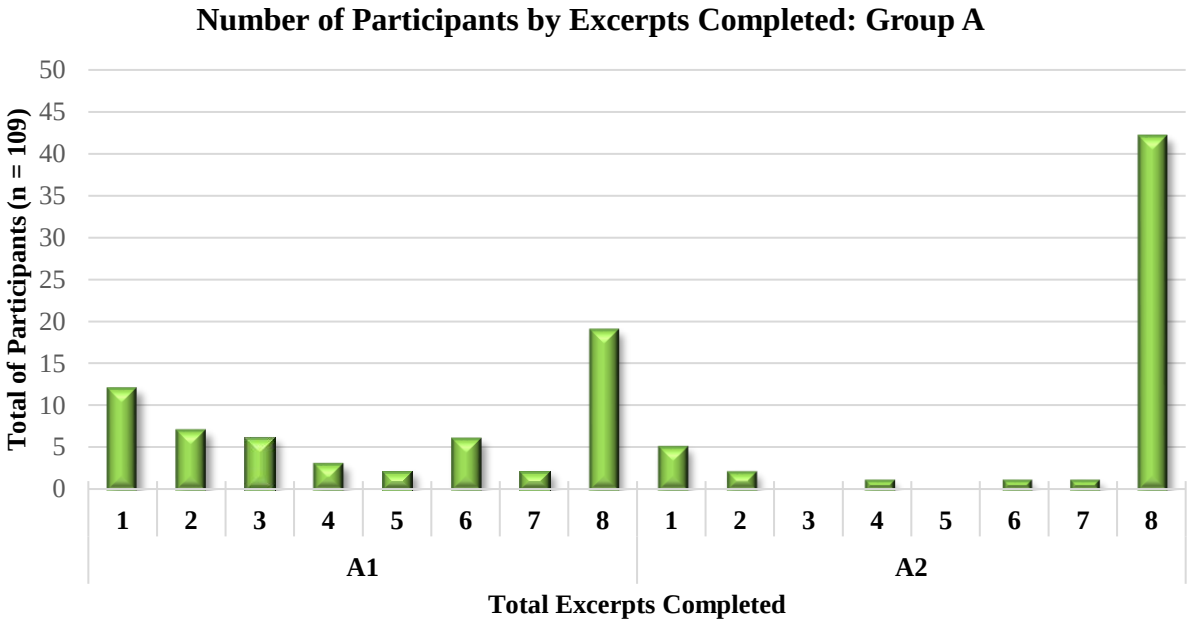
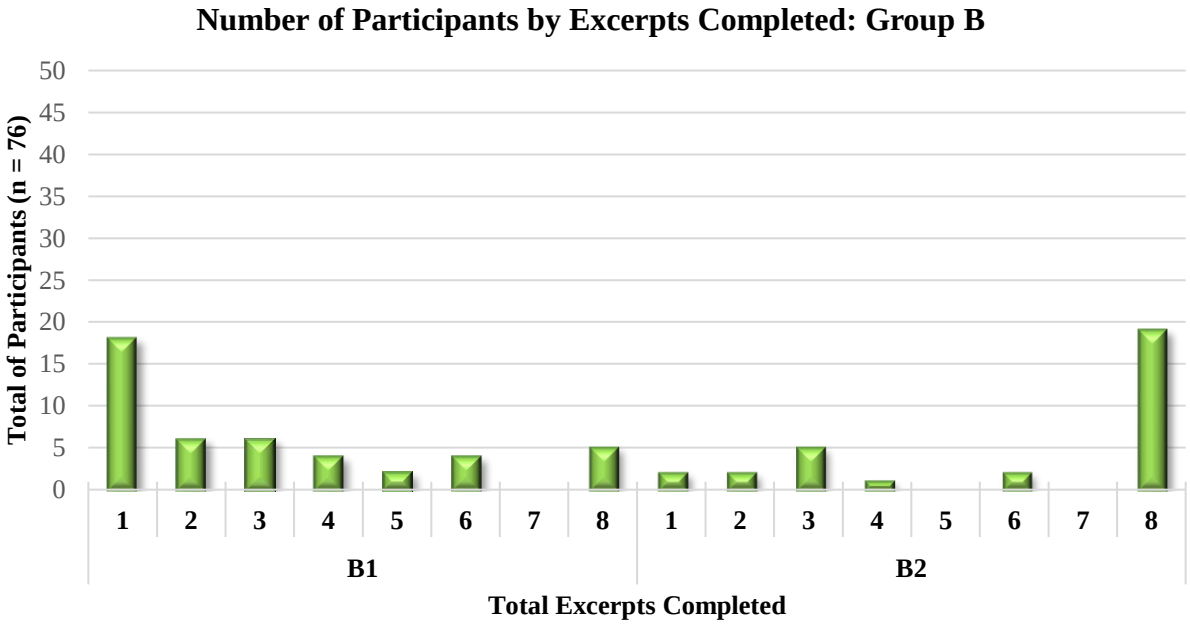


Figure 4.15.



In Figures 4.14. and 4.15., the x-axis does *not* refer to an excerpt number, as they are numbered elsewhere, but rather is a count of the number of excerpts completed. As described earlier in the chapter and in Chapter 2, the excerpts were displayed in a random order, such that, for example,

the first excerpt viewed may or may not have been Excerpt 1. The y-axis is a count of the number of participants who responded to survey items on a given number of excerpts and then ceased participation in the survey at that point. Full completion entails completion of eight excerpts on two surveys, meaning that the full completion participants are illustrated for each group by the rightmost column in each figure.

For this survey, as described previously, participants who followed the survey link were presented with an informed consent page, followed by a demographics survey and an equipment test page. After clicking past the equipment test page, participants were without their knowledge randomly assigned to either Group A or B, and viewed a page of instructions indicating that they were now beginning the first of two surveys. Only those participants who then clicked through to an excerpt and completed at least one survey item pertaining to that excerpt were considered enrolled in the study and counted in the total number of participants; participants who only completed the demographics questionnaire but no survey items were excluded, as this did not constitute full enrollment in the study. After completing a total of eight excerpts, participants viewed a page which informed them that they had completed the first survey, and would now move to the second survey, with additional instructions about how to proceed.

Based on this design, it was anticipated that among those participants who enrolled in the study and completed at least one survey item, there would be three major points of attrition: after completing one excerpt (that is, out of curiosity to see the survey, but with no interest in full participation), at the halfway point after the first survey, and at the end of the second survey. For Group A, this anticipated pattern of attrition is borne out, with clearly visible spikes in the number of participants who quit the survey after one excerpt, after eight excerpts at the end of Survey A1, and at the end of Survey A2. In contrast, we only observe an uptick in participants

from Group B discontinuing the survey after one excerpt and at the end of Survey B2, with no discernible spike at the midway point at the end of Survey B1.

In addition to this somewhat unexpected pattern of attrition, Group B also exhibited a higher rate of attrition. The number of unique participants who completed at least one excerpt in Group A totals 109, of which nearly half completed the first survey and went on to complete at least one excerpt on Survey A2. Of the 52 participants who started Survey A2, over 80% (42) continued to full completion, which represents an overall completion rate of 38.5%. In Group B, there were a total of 76 unique participants who completed at least one excerpt. Approximately 40% (31) of the original participants completed the first survey and at least one excerpt on Survey B2. However, of the 31 participants who started the second survey, only 61% (19) completed both surveys in their entirety, representing an overall completion rate of 25% for Group B. In short, participants who were assigned to Group B were less likely to complete the survey overall as compared to Group A, and they were more likely to discontinue their participation in the midst of the second survey rather than at the more natural stopping point at the end of the first survey.

These differential patterns of survey attrition may offer some insight into the differences observed above in how all participants as compared to full completions performed on the second survey in Group A versus Group B. Recall that for both Group A and B, we find that the full completion participants tended to display higher accuracy and sensitivity in differentiating utterances treated as non-serious from those treated as serious. As illustrated in Figure 4.14., above, since most participants in Group A who initiated Survey A2 were also those who completed the entire survey set, it makes sense that we would see less change in accuracy rates and signal sensitivity on Survey A2 from all participants to full completions, as compared to

Group B, where nearly 40% of the participants who initiated Survey B2 did not then go on to complete the survey.

The reasons why Group B exhibited a somewhat unusual pattern of attrition and a higher rate of attrition compared to Group A remain unclear. One plausible reason for the differences in attrition patterns and rates may be that watching a video (as did Group A) is simply more entertaining than reading a transcript (as did Group B). Beyond this, watching videos is a normal daily activity for most people in contemporary Western society, while reading a transcript is a more unusual activity. Finally, watching videos of people engaged in conversation may entail a degree of inherent voyeurism which encourages engagement with the survey materials. Ultimately, all of this is mere speculation, but it may potentially warrant further investigation in future, especially with regard to best practices in survey design. It is noted here solely as a point of interest and as potentially explanatory for the somewhat larger change in accuracy on Survey B2 from all participants to full completion participants as compared to Survey A2 participants.

Despite the differences in how full completion participants performed on Survey A2 versus B2 as compared to all participants, the overall picture remains the same for both experimental groups, regardless of completion status. Specifically, both groups have increased accuracy on utterances treated as non-serious on Survey 2 compared to Survey 1, while simultaneously having slightly reduced accuracy on utterances treated as serious. In other words, as can be clearly seen in the change in accuracy rate portion of Figures 4.12. and 4.13., there is a convergence of accuracy rates on non-serious and serious utterances towards a level rate of accuracy on both utterance types. Because this overall trend is largely similar for both groups regardless of completion status, we will now focus our analysis solely on the full completion participants for the remainder of the chapter, as doing so has little effect on the trends

in the data while also offering the most complete data set on which to base our conclusions. All figures from this point forward in the chapter are therefore based on data from the full completion participants only. For the sake of comparison, the figures in Appendix 4.1. display data from all participants; one will find that they are indeed nearly identical.

Turning now to the full completion data in Figure 4.13., as noted above, there is a convergence of accuracy rates from Survey 1 to Survey 2 for both experimental groups. Participants tended to have increased accuracy on non-serious utterances on Survey 2, jumping from approximately 45% correct on Survey 1 in both Group A and B, to 63.7% on Survey A2 and 67.1% on Survey B2. Conversely, both groups tended to have reduced accuracy on serious utterances, dropping from over 75% on Survey A1 to approximately 65% on Survey A2, and from over 80% on Survey B1 to approximately 70% on Survey B2. Despite the decreased accuracy on serious utterances, in combination with the increased accuracy on non-serious utterances, both groups displayed overall increased sensitivity to non-seriousness on the second survey (A' = 0.722 (A2:FC) and 0.769 (B2:FC)) as compared to the first (A' = 0.684 (A1:FC) and 0.735 (B1:FC)), meaning they were more readily able to differentiate between non-serious and serious utterances on the second survey compared to the first.

Furthermore, both Group A and B had improved response bias on Survey 2 as compared to Survey 1. Both groups had a response bias towards "No" on Survey 1 (B'' = 0.146 (A1:FC) and 0.246 (B1:FC)), but this was reduced to nearly no response bias on Survey 2 (B'' = 0.007 (A2:FC) and 0.022 (B2:FC)). In conjunction with the increased overall sensitivity to non-seriousness on Survey 2, this decrease in response bias offers some initial evidence in support of the view that [- humor] is the default, unmarked form of the [+/- humor] set. In other words, in the absence of clear evidence to the contrary, participants seem biased towards concluding that a

given utterance is [- humor], especially on initial response. With more time to consider their response and additional evidence obtained by listening to the clips and/or reading the transcripts multiple times, this response bias seems to diminish. This finding dovetails nicely with the underlying premise explored in Chapter 3, namely that participants in conversation may mark an utterance as non-serious by manipulating a variety of acoustic features, whereby [+ humor] is the linguistically marked form.

As found in the previous section, Group B continues to display slightly higher rates of accuracy and correspondingly greater sensitivity in differentiating non-seriousness from seriousness than Group A, both overall and on Survey 1 versus Survey 2. Since these results hold for all participants as for full completion participants, the rest of the chapter will focus solely on the full completion participants. In addition, we found further evidence in support of the view of [- humor] as the default, unmarked linguistic form. We continue to explore the differences in performance between Group A and B by examining their results on each excerpt and how these correspond to the reasons stated for their responses, before turning to a brief consideration of the demographic survey.

4.4.4. Survey Results by Excerpt

To this point in the analysis, Group B has performed slightly better on both surveys in their experimental condition than did Group A. In addition, results have indicated that participants in the perceptual survey from both experimental groups tend to display increased sensitivity to non-seriousness, and therefore increased accuracy overall in differentiating non-serious from serious utterances, on Survey 2 compared to Survey 1. We now consider how results varied by excerpt.

According to the results of the sequential analyses presented in Chapter 3, these excerpts

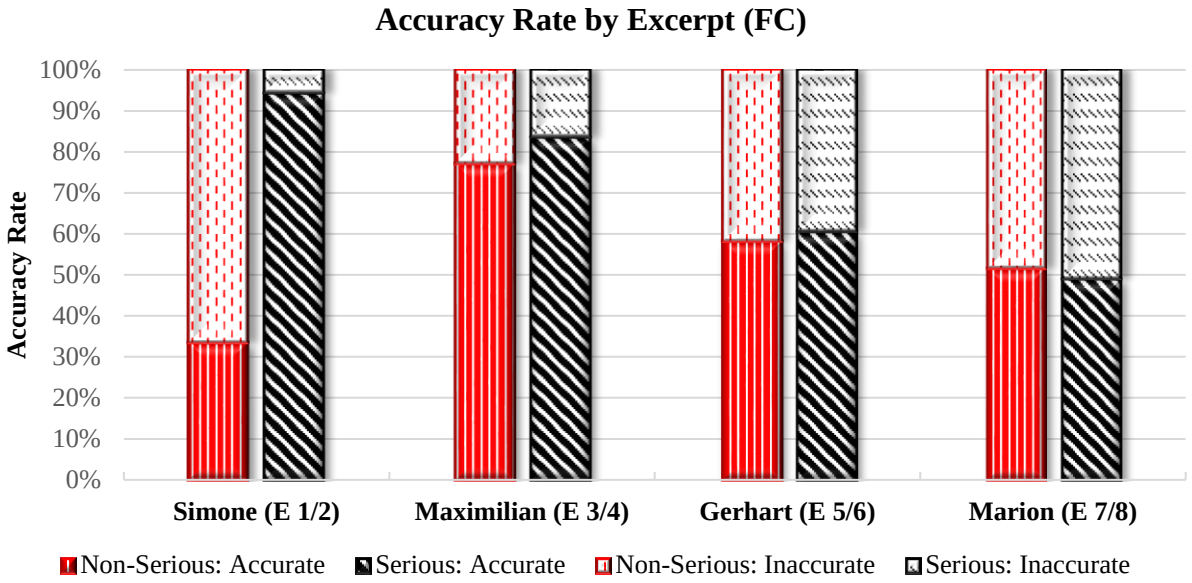
consisted of four utterances originally treated as non-serious and four treated as serious. There were four speakers in total, with one non-serious and one serious utterance from each speaker. All speakers spoke relatively standard varieties of colloquial High German, inflected with some regional variations in pronunciation and word choice. Two of the speakers spoke with regional hallmarks of Hessian, and two with hallmarks of Berlin-Brandenburgish. There was one male and one female speaker of each regional variety of German, and all were native speakers of German. The speakers, with corresponding excerpt numbers, are summarized in Table 4.16., below.

Table 4.16.

Speaker Demographics		
Speaker	**Gender**	**Regional Variety**
Simone (E 1/2)	female	Berlin-Brandenburgish
Maximilian (E 3/4)	male	Hessian
Gerhart (E 5/6)	male	Berlin-Brandenburgish
Marion (E 7/8)	female	Hessian

Figure 4.17., below, contains a summary of overall accuracy rates for each speaker, with accuracy rates for non-serious utterances (Excerpts 1, 3, 5, and 7) indicated to the left, and for serious utterances (Excerpts 2, 4, 6, and 8) to the right of each pair of bars. These overall results are the combined accuracy rates of the full completion participants from Group A and B on both surveys.

Figure 4.17.

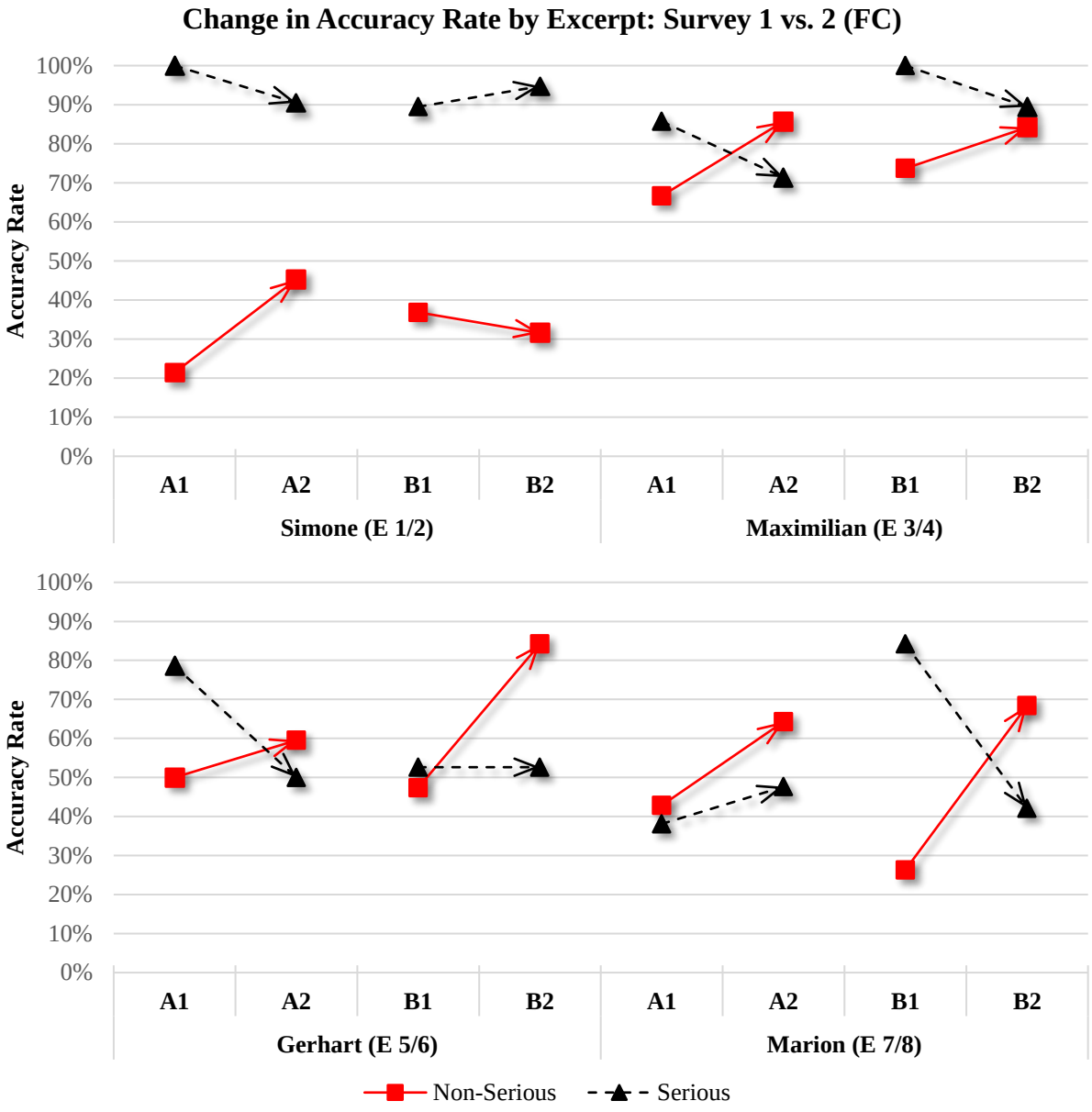


On initial inspection, it seems clear that participants were most successful at differentiating the non-serious utterances from their serious counterparts with the speakers Maximilian and Gerhart. With Marion, they were less successful, hovering near 50% accuracy on both her non-serious and serious utterances. Participants had the lowest rate of accuracy on Simone's non-serious utterance, at approximately 34% overall, while they identified her serious utterance with a nearly 95% rate of accuracy.

These observations offer the first impression that Maximilian and Gerhart were our most "successful" speakers of non-seriousness, in that survey participants were more likely to accurately identify how their utterances were treated in the original conversation and in turn the utterances' [+/- humor] status. By extension, Simone was least successful, while Marion was somewhere in the middle. However, closer inspection of how survey participants performed in Group A versus B, and on Survey 1 versus 2, reveals that this interpretation of the overall accuracy rates is overly simplistic. Figure 4.18., below, compares how survey participants' accuracy rates differed between the experimental conditions for Group A and B, and also

illustrates how these changed from Survey 1 to 2 within the respective groups.

Figure 4.18.



Comparing the results from Group A versus B, we find that dividing these out by speaker complicates the overall picture that has emerged heretofore suggesting that Group B generally has higher accuracy than Group A. Of particular interest are the results on the first half of the survey, where the survey conditions differ most radically (with Survey A1 offering video clips of the excerpts with no transcripts, and Survey B1 transcripts only). Here, we find that Group B

participants indeed have a higher rate of accuracy with the non-serious excerpts from speakers Simone (Excerpt 1) and Maximilian (Excerpt 3), but not with Gerhart (Excerpt 5) or Marion (Excerpt 7), where Group A participants perform slightly better on the first survey. Similarly, Group B has a higher rate of accuracy on only two of the four serious utterances from the first survey, outperforming Group A on the serious excerpts from Maximilian (Excerpt 4) and Marion (Excerpt 8).

These results taken solely in terms of accuracy rates present a somewhat confusing picture; as in previous sections, consideration of the data through the lens of signal detection theory (Table 4.19., below) again offers some additional insights.

Table 4.19.

Signal Detection and Response Bias by Speaker (FC)			
Speaker	**Survey**	***A′***	***B″***
Simone (E 1/2)	**A1**	0.804	0.988
	A2	0.796	0.484
	B1	0.752	0.424
	B2	0.778	0.625
Maximilian (E 3/4)	**A1**	0.849	0.289
	A2	0.867	-0.250
	B1	0.934	0.990
	B2	0.925	0.171
Gerhart (E 5/6)	**A1**	0.734	0.195
	A2	0.588	-0.018
	B1	0.500	0.000
	B2	0.784	-0.304
Marion (E 7/8)	**A1**	0.340	-0.019
	A2	0.609	-0.041
	B1	0.631	0.186
	B2	0.601	-0.060

On the first survey, Group A is more sensitive to non-seriousness from speakers Simone and Gerhart as compared to Group B. In the case of Simone, this result may seem counterintuitive, given the substantially higher rate of accuracy on non-serious utterances on Survey B1 (36.8%) compared to A1 (21.4%). However, this value for *A′* on Survey A1 is largely driven by the extremely high rate of accuracy on her serious utterance, which was at 100%, compared to 89.5% on Survey B1. Because *A′* as an index of sensitivity reflects the ability to identify

instances of the signal (or non-seriousness) accurately as well as to eliminate noise (or seriousness), it therefore stands to reason that participants on Survey A1 were indeed slightly more sensitive to the differences between the two simply by virtue of having not routinely mistaken instances of noise for signal.

The responses to Gerhart's utterances on the first survey are of particular interest here as well. On the transcript only condition for Group B, participants had a sensitivity index *A′* score of 0.500 (B1:E5/6), indicating no signal detection at all. In other words, participants could not distinguish between Gerhart's utterances treated as non-serious versus serious, such that they would have performed equally well by guessing at random without reading the transcript. In contrast, when viewing and listening to the excerpts on Survey A1, participants displayed greater sensitivity to non-seriousness, with an *A′* score of 0.734 (A1:E5/6). For speakers Simone and Gerhart, we conclude then that participants in the first perceptual survey were better able to differentiate between utterances treated as serious versus non-serious when they had access to the acoustic (and visual) signal.

In contrast, participants in the first survey displayed greater sensitivity to non-seriousness under the B condition for speakers Maximilian and Marion. Across the board, survey participants had high signal detection values with Maximilian's utterances, with Group B outperforming Group A slightly. On the other hand, Group A survey participants had a signal detection *A′* value of 0.340 (A1:E7/8) on the first survey when responding to Marion's excerpts, indicating that they were confusing the distractor (utterance treated as serious) for the signal (utterance treated as non-serious) and vice versa.

In terms of changes in signal detection from the first to the second survey, Group A participants had greater sensitivity to non-seriousness for Maximilian's and Marion's excerpts,

but not for Simone's and Gerhart's. The reverse is true for Group B participants, that is, they displayed greater signal sensitivity for Simone's and Gerhart's utterances on the second survey compared to the first, but not for Maximilian's and Marion's. However, only in the case of Gerhart and Marion (Group A only) were these changes in signal detection especially large, which explains why both experimental groups in aggregate had greater signal sensitivity on Survey 2 as compared to Survey 1, despite the slight variation in results by speaker.

Turning now to response bias, recall that in Section 4.4.3., participants in both survey Groups A and B were generally biased towards a "No, not a joke" response on both surveys, however, both groups also were less biased towards a "No" response on the second survey as compared to the first. This, I argued, supported the evidence in Chapter 3 that [- humor] status is treated as the default, unmarked assumption for a given utterance. A similar pattern holds when examining the data by speaker: in both Groups A and B, participants were less biased towards a "No" response on the second survey as compared to the first (Table 4.19., above). The sole exception to this was survey participants in Group B responding to Simone (Excerpts 1/2), who tended to be more biased towards a "No" response on the second survey.

Also of note is that Group A participants had a negative *B''* value on the second survey when responding to Maximilian and Marion, indicating a bias towards a "Yes, is a joke" response. It is possible that this was generally the result of a priming effect, as participants moved through the survey and perhaps began to understand what the subject of investigation was (i.e., perception of non-seriousness). However, in the case of Marion in particular, we find that participants in Group A already displayed a response bias towards a "Yes" response on the first survey, and furthermore that Group B participants had a similar bias for a "Yes" response on the second survey, which is not the case for any other speaker.

Overall, the results on the perceptual survey broken out by speaker followed many of the same trends found in the preceding sections, in that participants generally had greater accuracy in identifying utterances treated as non-serious on the second survey, with lower accuracy with the serious utterances. Although this was not the case for all speakers, the overall trend towards converging accuracy rates is visible in Figure 4.18., with the exception of Marion for Group A and Simone for Group B. In terms of signal detection, participants were also able to differentiate between signal (non-seriousness) and noise (seriousness) from most speakers, with similar results compared to those found in Section 4.4.3.

Of note is that survey participants had quite low accuracy rates with Simone's non-serious excerpt, yet had higher accuracy rates with Marion's non-serious utterance. This was somewhat surprising, given the results presented in Chapter 3, where it was found that Marion's non-serious and serious utterances were acoustic outliers as compared to the other speakers, so it was thought that survey participants might struggle most with her non-serious utterance. However, despite the low accuracy rate on Simone's non-serious utterance, survey participants' *A′* values indicate that they were able to differentiate between her serious and non-serious utterances, meaning that the lowered accuracy rate with the non-serious utterance is potentially attributable to an especially strong response bias towards a "No" response with this speaker (as indicated by the high *B″* values). In contrast, the *A′* values associated with Marion's excerpts, especially on Survey A1, indicate signal confusion, while her *B″* values reveal bias towards a "Yes" response. As such, despite the lower accuracy rates associated with Simone's non-serious utterance, there is some initial evidence from a signal detection perspective that, in fact, participants were more successful in predicting whether her utterances were treated as non-serious versus serious than they were with Marion's after all.

4.4.5. Survey Results by Reasons Given for Responses

In the preceding sections, we viewed the results for each experimental group over the course of the survey set and broken down according to the speakers of each excerpt. These results indicated that participants in the survey were, by and large, able to identify when a non-serious utterance was made, and therefore to predict when laughter was likely to follow such an utterance, with a rate of accuracy above what chance would predict. We turn now to consideration of the reasons participants gave for their assessment of a given utterance as serious or non-serious.

Participants' answers to the survey prompts were first coded and bundled as simple "Yes, a joke was made," or "No, a joke was not made," responses, following the procedure described above in Section 4.4.1. As illustrated above in Table 4.3. and Figure 4.4., participants in Group A were then asked only on Survey 2 to state the reasons why they responded as they did to each excerpt, whereas participants in Group B were asked to state the reasoning behind their responses on both Survey 1 and Survey 2. An explanation for this survey design can be found in Chapter 2, and the full text for the prompts can be viewed in Appendix 2.4.

The reasons participants gave for why they responded as they did to the survey excerpts were collected using open text boxes with no character limit or other restrictions on what could be entered. Given the nature of open text responses, there was naturally substantial variation in the exact phrasing of participants' responses, and the sorts of reasons they stated. Despite this, the responses could easily be grouped into categories of similar issues or topics addressed in the text participants submitted. These categories were determined based on how participants themselves responded to the prompts. In other words, the categories were not selected until the data had already been collected. This tagging procedure was necessary in order to quantify and

compare free text entries across survey participants; without aggregating and tagging responses, this is otherwise impossible.

All free text responses were therefore tagged as belonging to one or more of the categories found in Table 4.20. below

Table 4.20.

Response Categories for Reasons Given	
Response Tags	**Definition**
Body Language	speakers' body positions, gestures, facial expressions
Content	words used in utterance, phrasing
Context	surrounding conversation and/or prior utterances
Environment	physical appearance of room, items in view
Pragmatics	interactional behaviors described as normal, expected
Sound	tone of voice, how utterance sounds
Transcript	references transcripts, i.e., punctuation, spelling, etc.
Other	does not fit with other tags, often unusual responses
Unsure / No Response	statements of not knowing or being unsure, and/or left blank

As stated in Section 4.4.2., the responses to each excerpt were first bundled as a set, and coded as either "Yes, a joke was made," or "No, a joke was not made." All text entries were similarly bundled. Given that survey participants would often give more than one reason for why they selected the responses they did to the first two questions about what had just happened and what would happen next, the responses were frequently tagged with multiple of the above tags.

The selection and definition of the tags was guided entirely by how participants responded to the survey items; as stated above, the categories were not predetermined, but were selected based on natural categories that emerged from participants' responses. Furthermore, these tags should not be understood as referencing technical, linguistic jargon. On the contrary,

the tags reference a layman's understanding of the categories, as illustrated in Table 4.20. above and with the example responses described below.

The category of Body Language was selected if participants made any sort of reference to bodily movements, gestures, and/or facial expressions. This includes a more technical understanding of body language, but also any sort of description of bodily actions, even if such descriptions might not normally be considered an example of "body language" as such. Many participants named body language specifically, using not only the term *Körpersprache* (body language), but also closely related terms such as *Körperhaltung* (posture), *Gestik* (gestures), *Mimik* (facial expressions), and *Blickrichtung* (direction of gaze). Some participants simply described the bodily movements they saw in the videos, in statements such as the following:

Table 4.21.

Example Responses: Body Language	
Original German	**English Translation**
Händespiel, nach unten sehen	fiddling with hands, looking down
Sprecherin senkt am Ende den Kopf	speaker lowers head at the end
Kopfbewegung, Gesichtsausdruck	head movement, facial expression
[Teilnehmer] E lächelt	[Participant] E smiles
[Teilnehmer] E hört zu, sieht zustimmend aus	[Participant] E listens, appears to agree
aufrecht, aber nicht versteift	upright, but not stiff
Schulterzucken am Ende	shoulder shrug at the end
an die Nase fassen	touching nose
über die Stirn reib[en]	rubbing over forehead
an den Kopf fassen	touching head

Such responses clearly reference bodily movements on the part of the speakers seen in the videos, without using any of the more technical descriptors referenced above. Any such

descriptions were therefore included in the Body Language category, along with more direct references to body language and gesture. It should be noted that this category could only be selected for responses on Surveys A2 and B2, as these were the only free response survey segments which included video. Any responses on Survey B1 which seemed to indicate body language were assigned to the Other category, as it was unclear how participants could consider body language when presented only with a transcript.

The categories Content and Context occurred with a great deal of overlap, and are therefore best understood in terms of how they relate to each other. Content refers to what was said *in the utterance* in question, which was marked with arrows in the transcripts provided to survey participants. Context, on the other hand, refers to anything that was said in the excerpt *prior to* the utterance of interest, i.e., all other utterances in the excerpt which were *not* marked with arrows in the transcripts. Participants especially attuned to the exact words and/or phrases used in the excerpts often made reference to both the words in the utterance itself, and also to previous statements in the excerpt, with the result that their responses were frequently tagged with both of these categories. As with the Body Language category, these two categories were often named explicitly. Responses such as *Inhalt* (content), *Wortwahl* (word choice), *Thema* (topic), and *Semantik* (semantics) were all tagged as one or both of the categories Content and Context. In cases where participants provided one of these responses without any clarification as to which portion of the excerpt was being referenced (i.e., the utterance proper, or the preceding context), then both categories were selected.

In addition to these explicit statements referencing conversational content and/or context, participants often either summarized the conversation or quoted some portion of the excerpt. Summaries and quotes that referenced only the utterance itself were tagged as Content, while

those that referenced the preceding conversation were tagged as Context. If summaries and quotes referenced both the utterance itself and the preceding conversation, both categories were selected. Examples of utterances marked as Content, Context, or both are given in Table 4.22.

Table 4.22.

Example Responses: Content and Context	
Original German	**English Translation**
[Teilnehmer] G kritisiert ein Informatikstudium	[Participant] G criticizes majoring in computer science
ein lustiger Vergleich wurde gemacht	a funny comparison was made
„wie uncool“	“how uncool”
von Steine sammeln zu Versteinerungen	from collecting rocks to fossils
[Teilnehmer] M redet über Fußball	[Participant] M talks about soccer
das „nicht einschätzen können“	the part “can’t really evaluate”
„nee das ist ganz weit weg“	“no that’s really far away”
[Teilnehmer] M beschwert sich über etwas	[Participant] M complains about something
„bis zum Erbrechen“	“to the point of vomiting”
wiederholtes Nachfragen Ms	the repeated questions from M

Any summaries and/or quotes similar to the examples provided in Table 4.22., above, were included in one or both of the categories Content and Context, depending on which portion of the excerpt was summarized or quoted.

The category Environment does *not* refer to “environment” as the term is more typically used in linguistics to reference conversational environment; this (linguistic) definition of the term “environment” was instead included in the category Context, as described above. In contrast, the category Environment refers to any and all physical surroundings which are visible in the field of view of the camera. This category refers solely to any objects or items which were visible in the video excerpts, but which were not attributable to participants’ body language. In other words,

participants' outward appearance, such as descriptions of hair or clothing, could potentially fall under Environment, whilst playing with hair or buttons on a shirt, for example, would fall under Body Language. Survey participants occasionally referenced the physical surroundings explicitly, with responses such as *Atmosphäre* (atmosphere) or *Umgebung* (surroundings). However, as with the previous categories, they often described the environment instead, as in the following examples:

Table 4.23.

Example Responses: Environment	
Original German	**English Translation**
Weingläser auf dem Tisch	wine glasses on the table
beide [Teilnehmer] sehen gleich aus	both [participants] look the same
Bier steht auf dem Tisch	beer is on the table
es wird getrunken	[alcohol] is being drunk
Ähnlichkeit der Gesprächspartner	the similarity of the conversational partners
die Kumpel haben sich ein Bier aufgemacht	the pals have opened up a beer
weil es lustig ist, in einem rosa Zimmer…	because it's funny in a pink room…
Beide sitzen an einem Schreibtisch in einem WG-Zimmer	the pair is sitting at a desk in a shared apartment
Kleidung, Einrichtung, Weißwein	clothing, furnishings, white wine
Atmosphäre in der Küche, Bierflaschen	atmosphere in the kitchen, beer bottles

As is clear in the preceding examples, the category Environment was selected only in instances where survey participants explicitly described some feature of the room or items in the room that were visible in the video excerpts. Similar to the Body Language tag, this category could only be selected for responses on Surveys A2, and B2, i.e., the two open response survey segments which included video.

In addition to describing what they saw (body language, environment) and heard (content, context), survey participants frequently stated that they responded as they did to the survey questions because their responses to and/or predictions about the excerpts were logical, expected, or natural. In other words, they were explicitly drawing on a reservoir of interactional and cultural knowledge which gave them some sense, or innate feeling, about how a normal conversation *ought to* proceed under normal circumstances. In such cases, the tag Pragmatics was selected. A few survey participants actually entered *Pragmatik* (pragmatics) as part of their responses, while other participants entered related linguistic terms, such as *discourse marker* (sic) and *hedge* (sic). Most responses tagged in this category, however, did not include such technical terms, as can be seen in the examples below.

Table 4.24.

Example Responses: Pragmatics	
Original German	**English Translation**
weil 80% Ballbesitz und 2 Torschüsse einen krassen Widerspruch darstellen	because 80% ball possession and two shots at the goal stand in glaring contradiction
Diamant - unwahrscheinlich	a diamond - unlikely
typische Themeneinführung mit „auch“	typical introduction of a topic with “*auch*”
Stammtischschilder sind toll!	*Stammtisch* signs are awesome!
logischer Verlauf des Gesprächs	logical continuation of the conversation
über Fußball können Männer immer viel reden	men can always talk a lot about soccer
Fußballgespräche werden immer fortgeführt… ;-)	conversations about soccer always continue on some more… ;-)
Gemeckert wird immer gerne	people like to complain
weil ein Stammtischschild extrem altmodisch ist und nur in traditionellen Gasthäusern vorkommt	because a *Stammtisch* sign is extremely old-fashioned and is only found in traditional restaurants
weil die Aussage so unlogisch ist	because the statement is so illogical

As can be seen in the above examples, terms such as *immer* (always), *logisch* (logical), and *typisch* (typical) indicate that these are simply expected, normal outcomes of the ongoing interaction. Even when such terms are not present, as in the example *Diamant - unwahrscheinlich* (a diamond - unlikely), the participants' responses make clear that these are entirely obvious, logical conclusions that any pragmatically competent speaker of the German language ought to understand.

In addition to the preceding categories, many survey participants specifically indicated that they had taken the sound of a given utterance under consideration while responding to the survey questions. The next tag, Sound, was selected for any survey participant responses which addressed *how* participants in the conversations sounded as they spoke, as opposed to *what* they said or the context in which it was said. Some examples of generic responses indicating attention to how participants spoke include entries such as *Ton* (sound), *Stimme* (voice), and *Aussprache* (pronunciation). Some survey participants also referenced more specific components of vocal production, with responses like *Tonfall* or *Intonation* (intonation), *Tonlage* (pitch), *Satzmelodie* or *Prosodie* (prosody), *Betonung* (stress, emphasis), and *Sprechlautstärke* (speaking volume). Also included in this category were myriad means of referring to rate of speech, including *Tempo* (tempo), *Sprechgeschwindigkeit* (rate of speaking), and *Pausen* (pauses). Aside from these explicit, but nonspecific, references to how speakers sounded, survey participants often described the excerpts' sound in rich detail, as found in the examples below.

Table 4.25.

Example Responses: Sound	
Original German	**English Translation**
Timing (ins Wort fallen)	timing (speaking over each other)
die Art und Weise, wie es gesagt wurde, zeigt dass es scherzhaft gemeint ist	the manner in which it was said shows that it is meant in jest
aus der Intonation ist ganz klar, dass das ein Vorwand ist, weil sie die Ausstellung nicht gut findet	from the intonation it's obvious that this is a pretense, because she doesn't think the exhibit is good
Lachen liegt schon in der Stimme	there is already laughter in the voice
auch wie er es sagt - der Tonfall ist nicht aggressiv, sondern den Trainer und die Situation ein bisschen lächerlich machend	also how he says it - the intonation is not aggressive, but rather pokes fun at the coach and the situation a bit
erst Kopfschütteln, dann Tonfall bei der Ablehnung; [Teilnehmer] M ist froh, dass sie nicht zur Ausstellung gehen „kann"	first head shaking, then intonation at the declination [of invitation]; [Participant] M is happy that she is not "able" to go to the exhibit
paraverbale Sprachebene: Senkung und Hebung der Stimme als Mittel zur Ironie / des Wortwitzes	paraverbal level of speech: lowering and raising voice as a means for doing irony / puns
[Teilnehmer] G redet schneller bei Witz	[Participant] G speaks faster during joke
steigende Intonationskurven	rising intonation curves
Betonung „und zwei" - unterstreicht den Gegensatz	stress on "and two" - underscores the contrast

Such descriptions of prosody, inflection, and other facets of vocal production clearly illustrate that survey participants were orienting to the sound of the speakers' voices. It should be noted that this category could only be selected for responses on Surveys A2, and B2, as these were the only two free response survey segments which included audio. Any responses on Survey B1 which seemed to indicate how something was said, or how it sounded (as in, "*klingt nach*..." ("sounds like…")), were assigned to the Content and/or Context category.

Furthermore, the Sound tag was only selected in instances where survey participants gave responses which contained keywords such as the generic terms listed above, or were descriptions

similar to the responses in Table 4.25. In cases where participants gave responses such as "*wie es gesagt wird*" ("how it is said") with no further clarification, given that *wie* (how) can be understood as referring to myriad aspects of speech, such as word choice, content, conversational context, and/or pragmatic considerations, as well as the sound of a given utterance, the tags Content and Context were selected as default variants of "how" something is said. As such, it is likely that there is a slight undercount of instances where Sound would have been appropriate, and a corresponding overcount of the categories Content and Context. However, in consideration of the fact that *most* responses clearly indicated Content and/or Context, as well as the potential effects of confirmation bias (in this case, with the working hypothesis that sound plays a key role in production and perception of non-seriousness), the decision was made that responses containing some variation of "how it was said", without any further clarification, would best be classified as Content and/or Context. This better suited the overall trends in the data, and also served to reduce the potential impact of confirmation bias.

The final named category, Transcript, was also the final category tag added. As responses were tagged, if the category Other grew especially large, the responses were reexamined to determine if there was any unifying feature which might be appropriately added as a named response tag. Both Environment and Transcript were added in this manner. On the other hand, the categories Body Language, Content, Context, Pragmatics, and Sound had all been determined in similar fashion using survey participants' response data in the pilot study, and were carried over for the full study. The answer tag Transcript was selected in cases where participants specifically referenced the transcript they read in their response. This was commonly stated explicitly as *Transkript* (transcript) or *Transkription* (transcription), however, participants also often describe the contents of the transcript as well, as in Table 4.26. below.

Table 4.26.

Example Responses: Transcript	
Original German	**English Translation**
Fehler (Grammatik)	mistake (grammar)
Gespräch ist wesentlich einfacher zu verstehen als das Transkript	conversation is substantially easier to understand than the transcript
siehe Transkription ("lacht")	see transcript ("laughs")
wegen des Fragezeichens	because of the question mark
Fragezeichen	question mark
die Pause ist viel kürzer als im Transkript angenommen; ich hätte gedacht, jeder Punkt steht für eine Sekunde, aber in Wirklichkeit ist die Pause quasi nichtexistent	the pause is much shorter than suggested by the transcript; I would have thought that each period indicates one second, but in reality the pause is basically nonexistent
aufgrund der Satzzeichen	because of the punctuation
Interpunktion im Transkript	punctuation in the transcript
Punktuation	punctuation
wahrscheinlich Frageintonation, verschriftet mit '?'	probably question intonation, indicated with '?'

The references to punctuation, in particular, indicate that survey participants in these instances were orienting to the what they saw in the transcript, in addition to, or as opposed to, what they heard or saw in the video. In the case of the response *Fehler (Grammatik)* (mistake (grammar)), the tag Transcript was selected because it was in response to Excerpt 7. Unfortunately, there was a small typo (consisting of a single, vital letter) in the transcript for Excerpt 7 in the live version of the survey. As such, any responses to Excerpt 7 which made note of any sort of grammatical "mistake" were taken as references to the typo in the transcript, and were tagged with this category.

The next category, Other, was selected as infrequently as possible, in order to avoid overuse as a catch-all category. As indicated above, responses in this category were occasionally

reassessed in order to determine if any new categories could be gleaned from the data, as was the case with Environment and Transcript, for example. However, there were also responses which were truly unique, or even at times strange. These were ultimately tagged as Other, examples of which are in Table 4.27., below.

Table 4.27.

Example Responses: Other	
Original German	**English Translation**
weil keine der anderen Antworten paßte	because none of the other responses fit
weil ich den nächsten Videoklip schon einmal gesehen habe	because I have already seen the next video clip
[ich] hoffe das ist ein Witz	[I] hope it is a joke
es gibt genug!	that's enough!
leider ist diese[s] Gespräch sehr gestellt / gekünstelt; die vielen Versprecher von [Teilnehmer] M weisen darauf hin, dass er versucht, sich einen Text zu merken, deshalb wirkt das Ganze nicht wirklich natürlich	unfortunately this conversation is very staged / artificial; the many slips of the tongue by [Participant] M show that he is trying to remember a script, so that the entire thing seems not very natural
Skins	skinheads
sie sind angetrunken	they are drunk
die Teilnehmerinnen haben gekifft	the participants smoked weed
Drogenkonsum führt nicht zwingend zu guten Unterhaltungen	drug use does not necessarily lead to good conversation
weil die Teilnehmer weiterhin bekifft sind	because the participants are still stoned

The responses in Table 4.27. are certainly fascinating in many regards. However, they do not form a cohesive set, or to the extent some of them do form a category, it was simply not large enough to be tagged across all survey participants. For example, while there were several variations on "none of the other answers fit", all of these responses were from a single participant (#178). The second response concerning having seen the next segment of video

seems to reference the participant's conclusions about how the survey was structured, as opposed to observations about context and/or pragmatics. The next two responses, in which participants state that they hope it is a joke and "that's enough!" seem to offer some insight into their feelings about the interactions, but not necessarily about what is happening in the excerpts. There were only a handful of these sorts of responses, so they were tagged as Other as well.

The next response, in which a participant indicates that the conversation seems staged due to "*die vielen Versprecher*" ("the many slips of the tongue") is a fascinating response in and of itself. This particular response lays bare one of the key challenges with using naturally occurring conversational data in a survey, namely, the general population is by and large unaware of how much the way we actually speak deviates from our memory of what is said in its idealized, sanitized form (Golato, 2003). This obviously poses unique challenges in survey design which must be considered in any future perceptual studies using naturally occurring data.

The final set of responses, unfortunately, seem to convey rather negative assessments of the character of the speakers in the excerpts. The assertion that the speakers are skinheads is understandable, if wholly inaccurate, based on the fact that the participants in Excerpts 5/6 had shaved their heads together as part of a bet on the day before the recording was made. The next response, indicating that the speakers were drunk, was also understandable, if inaccurate. Many respondents noted the presence of beer and/or wine in Excerpts 3/4 and 7/8; such responses were tagged Environment, as described above. However, this was the only response which specifically questioned the mental state of the participants, and was therefore tagged as Other.

In contrast, the final set of responses, in which the speakers are said to be "stoned" or "on drugs", has no basis in reality that I am able to ascertain in the video. The speakers had only recently woken up when they requested that I set up the camera, and they were still clothed in

pajamas and perhaps groggy, but they were not under the influence of any mind-altering substances, legal or otherwise. Given that these responses concerning illicit substance use, as well as some additional, similar ones, all came from a single survey participant, they were therefore also included, as outliers, in the Other category.

As a final note on this topic, inebriated individuals are *ipso facto* incapable of providing informed consent to participate in human research. As such, any potential participants who were known to have recently used any mind-altering substance, or were suspected of the same, were not permitted to consent to and/or participate in any recording sessions. Therefore, in order to obtain informed consent, all participants were sober at the time that any and all recordings commenced. Many participants did consume moderate amounts of alcohol during the recording process; this could not be easily prevented due to the study design, in which I informed participants that they could do whatever they wanted during the recording, as long as they generally limited outside noise (open windows, music playing in the background, etc.) and remained within the camera field. In short, all participants were sober when I arrived, when they signed their forms, and when I left the camera recording. No illegal substance use was recorded, nor would recordings of such use have been retained or distributed via survey. When I returned to collect the camera, I never observed any participants to be intoxicated. If they had been, I would have reviewed the recording(s) with the participant(s) when sober in order to reestablish the status of their consent for participation in the study; fortunately, this step was never necessary.

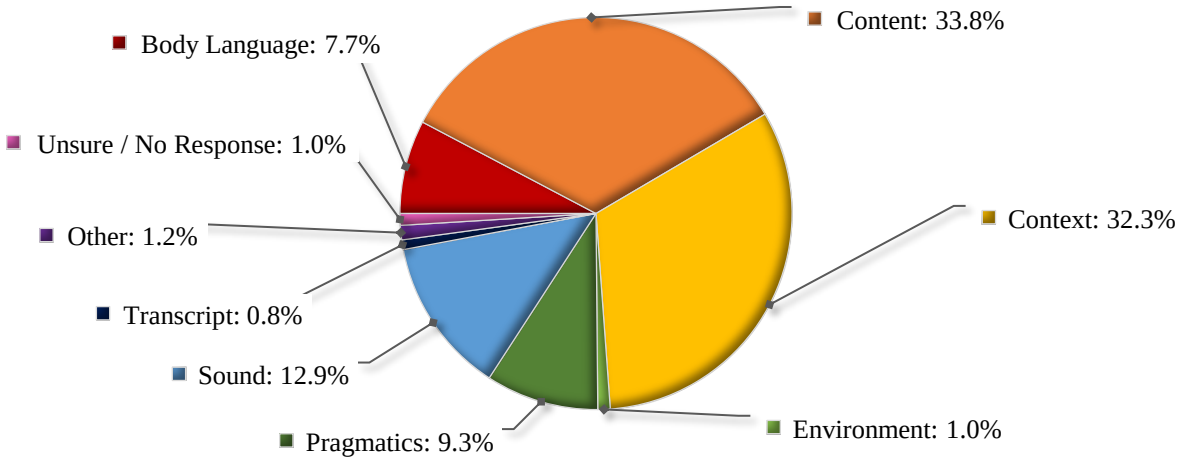
The final category, Unsure / No Response, was used for precisely two types of responses. First, if the text boxes were left entirely blank, or were filled with a nonsensical string of letters from keyboard mashing, the responses were tagged with this category. The second response type

tagged with this category was any variation on "*Ich weiß nicht*" ("I don't know"). Some survey participants stated that they didn't know why they responded as they did for half of a response bundle, but gave a more fulsome response to another portion of the excerpt; in such cases, it was possible for the response bundle to be included in the Unsure / No Response category in addition to other categories. Yet for the most part, survey participants who left some text boxes blank or explicitly stated that they did not know why they responded as they did tended to respond to all of the questions for a given excerpt in that manner. This was a rather small category, as the survey prompted participants to fill in any text boxes left empty when they clicked on *Weiter* (Next), although the survey did allow them to move on without prompt if they clicked on Next a second time.

Returning to the survey, we consider the results according to the tagged responses participants gave in the open text boxes on Surveys A2, B1, and B2. Overall, as can be seen in Figure 4.28. below, survey participants overwhelmingly indicated that they considered the Content and/or Context of the excerpts in giving their responses to the survey items.

Figure 4.28.

Reasons Given for Responses: Overall (FC)



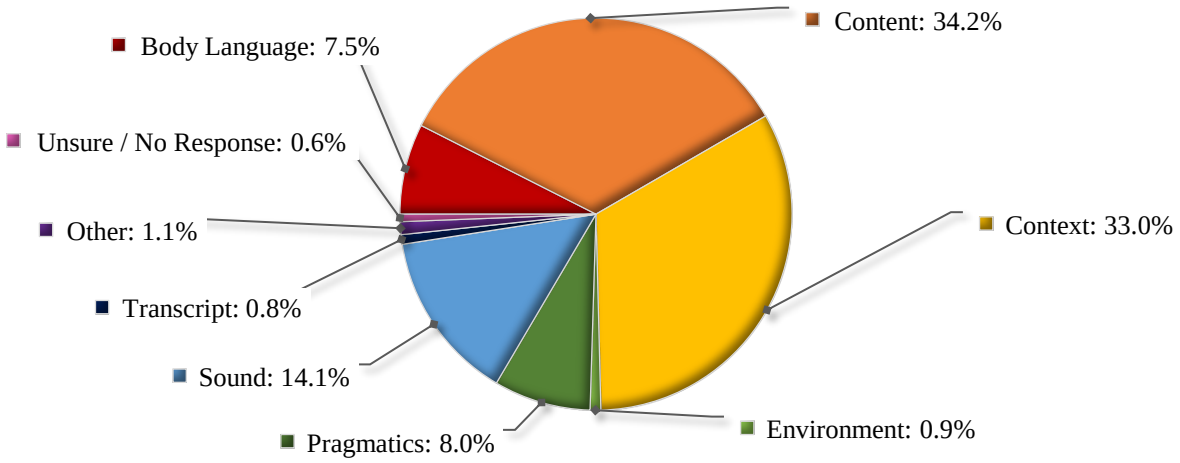
As is clear from the above, nearly two-thirds of the participants' responses were of the sort

tagged as Content and/or Context. This result is unsurprising, given that *what* is said in a given context is of tantamount importance in interaction.

However, if we consider only the responses which were "accurate", i.e., which identified the utterances treated as serious and non-serious by the original speakers, we find some interesting shifts in the proportional weights for each of the response tags, as in Figure 4.29.

Figure 4.29.

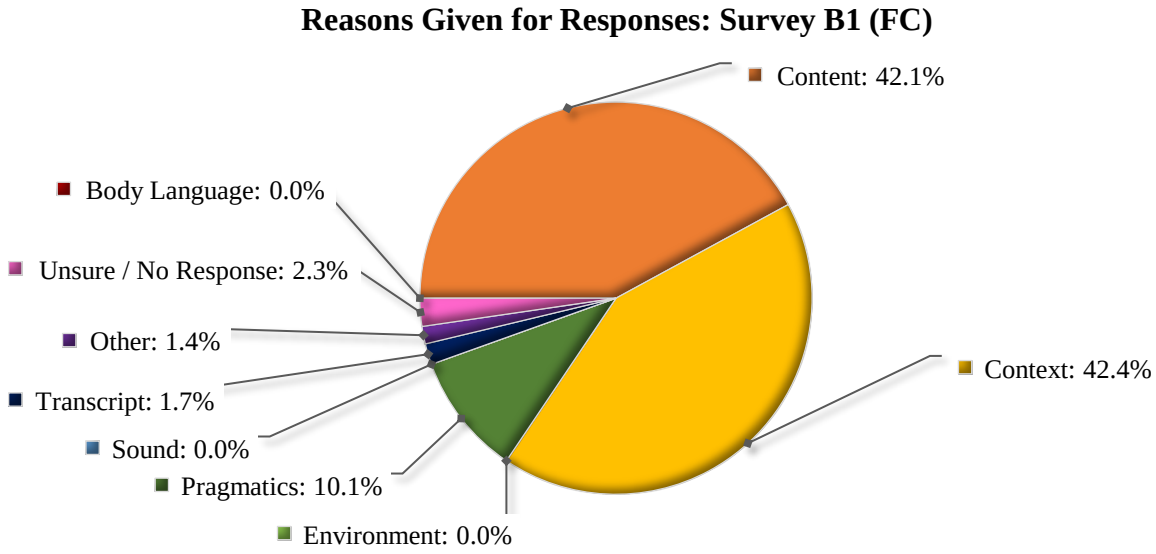
Reasons Given for Accurate Responses: Overall (FC)



For responses which accurately differentiated between utterances treated as serious versus non-serious, the categories Body Language, Content, Context, Environment, Transcript, and Other remain fairly stable. Responses tagged as Pragmatics and Unsure / No Response fall substantially, while the category Sound rises in corresponding fashion. This pattern, whereby a larger share of accurate responses (as defined in Section 4.3., above) were tagged with Sound-type responses, holds generally throughout the survey conditions where participants had access to the video, regardless of how the data are viewed.

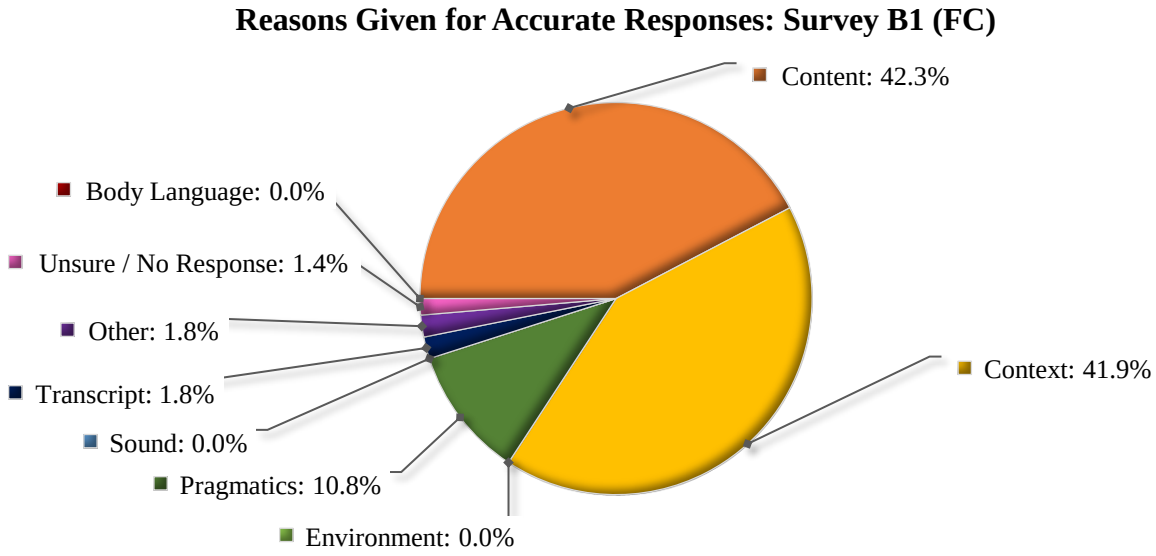
For example, bearing in mind that Survey B1 is the transcript only condition, it is unsurprising that Content and Context are the two largest categories, as in Figure 4.30.

Figure 4.30.



Indeed, these two categories comprise nearly 85% of the response types overall on Survey B1. Yet even if we only consider the accurate responses on Survey B1, there is very little movement between the response tags, as in Figure 4.31.

Figure 4.31.



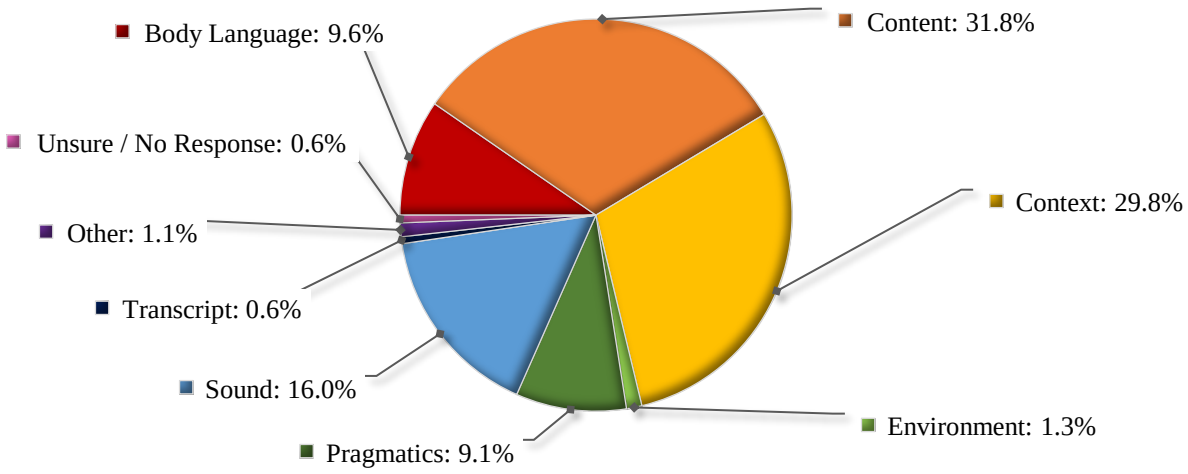
Here, we find that participants' accurate responses on Survey B1 shift slightly among the smaller categories, yet comparing accurate responses to the overall results in Figure 4.30., there is very little change in the general picture when participants had access only to the transcripts. The

greatest changes are in the Unsure / No Response category, which is much lower among accurate responses, and an increase in the share of responses associated with the Pragmatics tag. This stands in contrast to the results in Figures 4.28. and 4.29. above, where we see a decrease in responses tagged as Pragmatics when we consider only the accurate responses. It appears then that attending to Pragmatics is a useful strategy in perception of non-seriousness in the written word.

On the other hand, if we consider the results for Surveys A2 and B2 (the survey conditions with both video and transcript data; cf. Table 4.3. and Figure 4.4.), the general trends seen in Figures 4.28. and 4.29. above are again visible. As in Figure 4.32., Content and Context comprise again the largest share of the stated reasons for participants' survey responses.

Figure 4.32.

Reasons Given for Responses: Surveys A2 and B2 (FC)



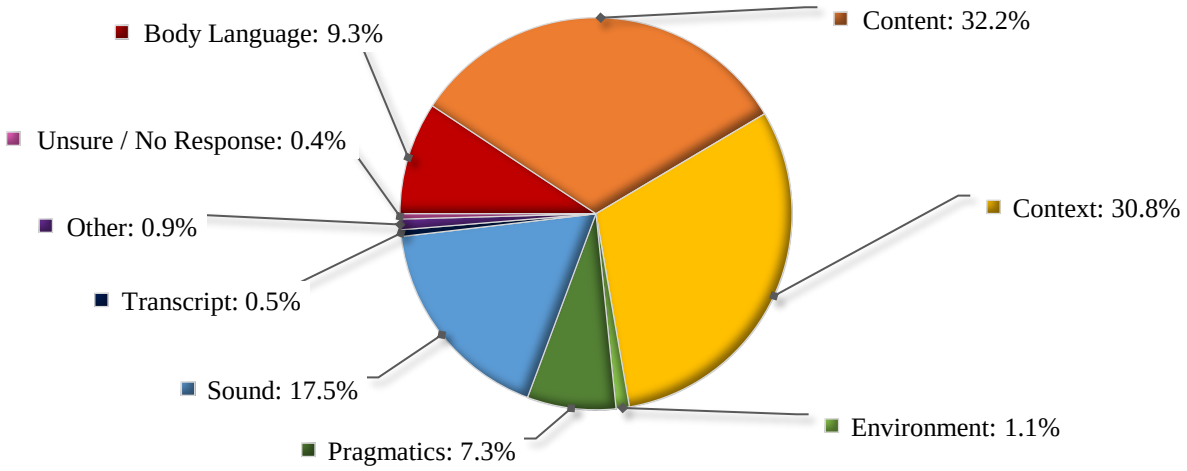
When the results for Survey B1 are excluded, the overall share for Content and Context is slightly reduced, from 66.1% in Figure 4.28. to approximately 61.5% in Figure 4.32., and the Unsure / No Response tags are also down from 1% to 0.6%. While Environment, Pragmatics, Transcript, and Other remain fairly steady, the two categories Body Language and Sound take up

a larger share. This result, obtained by eliminating Survey B1 from the overall results, is unsurprising, given that this considers only the two survey conditions with video.

Yet if we compare the overall results from Surveys A2 and B2 to the reasons given for accurate responses only on these two survey conditions, as in Figure 4.29., we again find the same trend as previously observed: a larger share of accurate responses indicated that participants had considered the sound of the utterance in responding to the survey items.

Figure 4.33.

Reasons Given for Accurate Responses: Surveys A2 and B2 (FC)



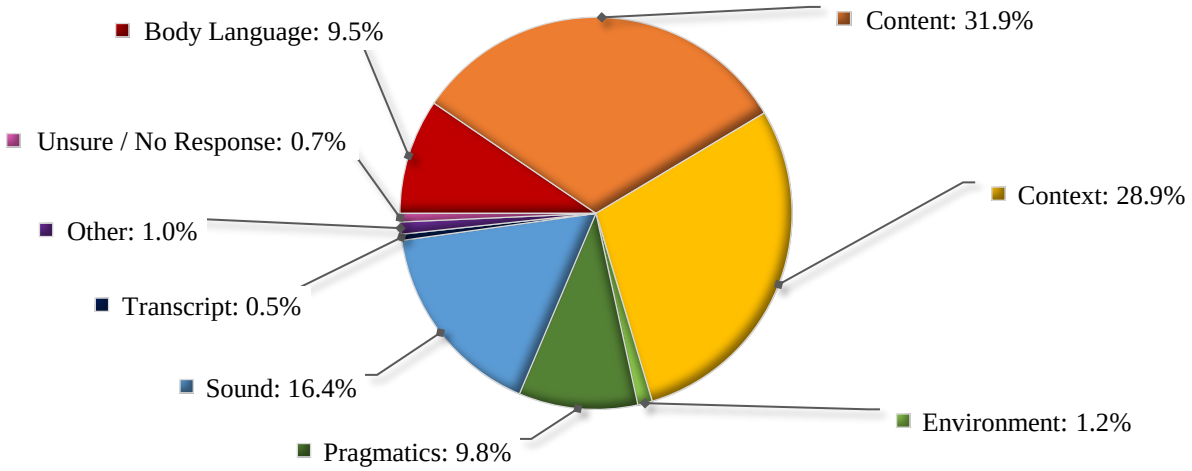
While most of the categories for accurate responses on Surveys A2 and B2 remain proportionally steady, there are slight reductions in all categories except Content and Context, which combined amount to 62% of participants' responses (compared to 61.5% in Figure 4.32., above). The reductions in all other categories, however, seem to have all shifted to Sound, where we find that 17.5% of participants' accurate responses, as compared to 16% overall, took the sound of the utterance into consideration. This again mirrors the overall results seen in Figures 4.28. and 4.29. above.

Unsurprisingly, if we consider Surveys A2 and B2 independently of each other, these same trends hold, i.e., a larger share of accurate responses indicate attention to Sound. Starting

with the overall results for Survey A2 in Figure 4.34., we first note that they are largely in line with the overall results for both surveys in Figure 4.32. above.

Figure 4.34.

Reasons Given for Responses: Survey A2 (FC)

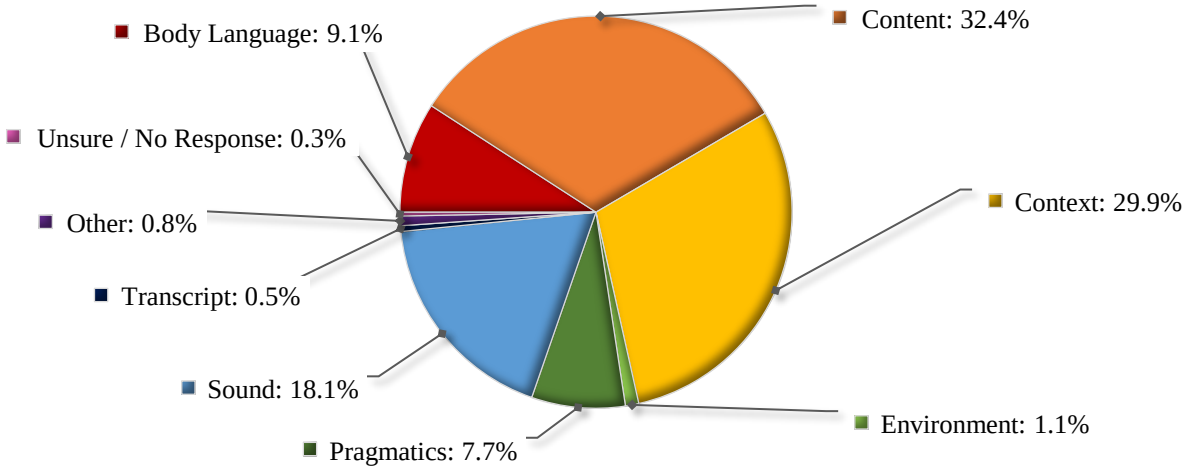


Perhaps most notable is that participants on Survey A2 indicated slightly lower attention to Content and Context (at 60.8% combined), with correspondingly increased attention to Pragmatics (at 9.8%); the other parameters remained fairly steady as compared to the overall results from both surveys combined in Figure 4.32. above.

Comparing the overall results to the accurate responses on Survey A2, we find again that there is a greater share of responses indicating attention to Sound, as seen in Figure 4.35. below.

Figure 4.35.

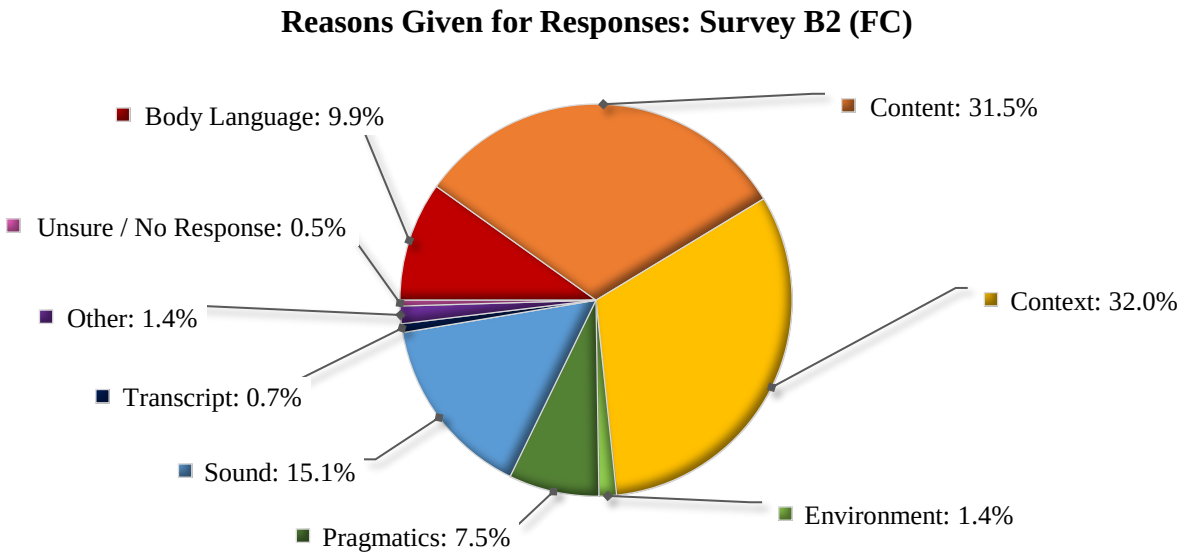
Reasons Given for Accurate Responses: Survey A2 (FC)



On Survey A2, we find that Content and Context combined, along with Body Language, Environment, Transcript, and Other remain fairly steady, while Pragmatics and Unsure / No Response comprise a reduced share of the responses. Most importantly, participants' stated consideration of Sound again represents a proportionally larger share of the accurate predictions that a given utterance would be treated as serious versus non-serious.

The results on Survey B2 are similar to those found on Survey A2. The overall response rates for each category tag are illustrated in Figure 4.36. below.

Figure 4.36.

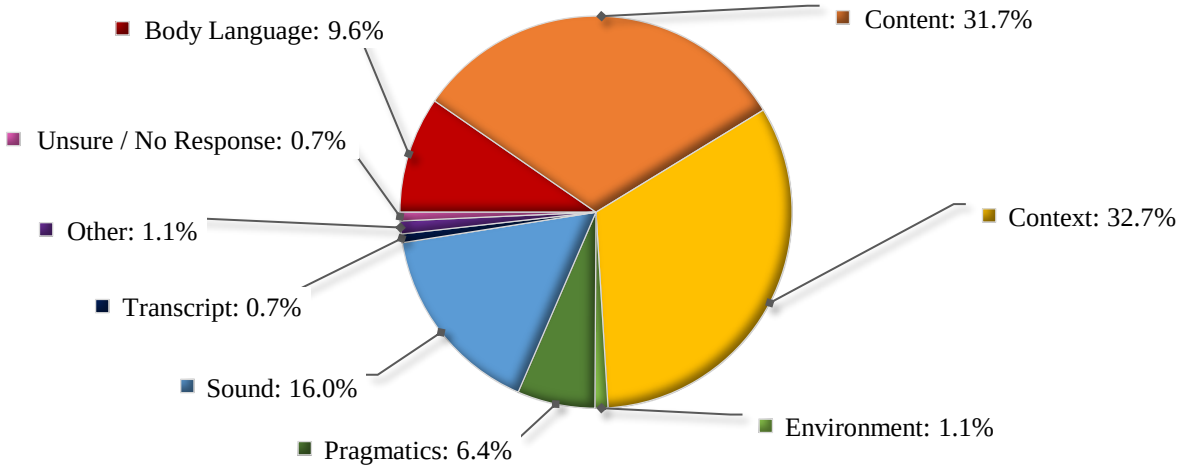


Mirroring the overall results seen for Survey A2 in Figure 4.34. above, Content and Context are again the most frequently stated reasons for participants' responses to the survey items. Most of the categories are similar in proportion to the overall results for both Surveys A2 and B2 combined in Figure 4.32. above, with the exception of Pragmatics and Sound, which are both lower on Survey B2 as compared to Survey A2.

Despite this, Sound still comprises a larger share of reasons given for accurate response on Survey B2, as seen in Figure 4.37. below.

Figure 4.37.

Reasons Given for Accurate Responses: Survey B2 (FC)



As with Survey A2, we find that participants report increased attention to Sound for accurate responses. All categories remain fairly stable for accurate responses, with the exception of Pragmatics, where participants reported less attention to this category on accurate predictions of whether utterances were treated as serious versus non-serious.

Returning to the reasons given for responses on Survey B1 (the transcript only condition, cf. Table 4.3. and Figure 4.4.), it seems clear that when participants only have access to transcripts, there is little change in the proportion of each category for overall results as compared to accurate results. The exception to this is the Pragmatics category, which has a slightly increased share among accurate predictions, even though this same category is reduced among the accurate responses on Surveys A2 and B2. This result offers some initial, tentative evidence in support of the hypothesis that the acoustic features of speech play an important role in the production and perception of non-seriousness, although it seems that attention to Pragmatics potentially played a role in perception of non-seriousness under the transcript only survey condition.

More compelling still are the results for Surveys A2 and B2. On these two surveys, we

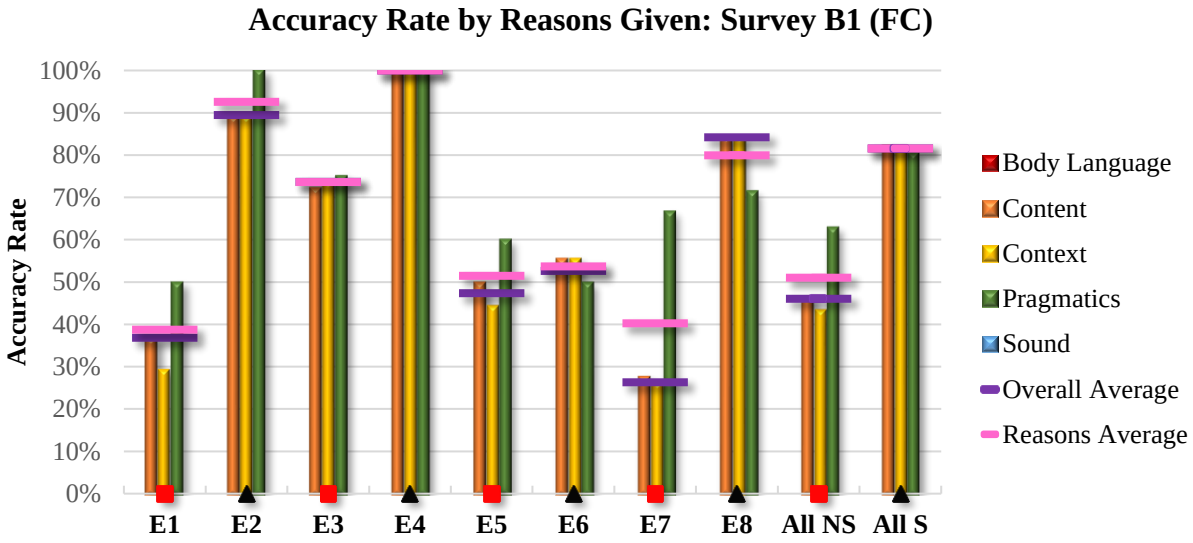
found that participants clearly reported attention to Sound, whether explicitly stating that they considered sound (*Ton*), or by offering detailed descriptions of various acoustic features of speech (cf. Table 4.25.). Furthermore, the Sound category occurs at a higher rate than the other categories when looking at accurate responses only. In other words, there are a greater share of survey responses for both Surveys A2 and B2 that accurately predicted whether a given utterance was treated as serious versus non-serious by the original speakers in the interaction, as compared to the responses associated with the Sound tag in the overall data. This means, in short, that listeners with access to impoverished conversational context that does not include any response from the original conversationalists, produce more accurate predictions about whether an utterance was treated as [+/- humor] when they claim to be attending to the acoustic aspects of the interaction.

The preceding analysis illustrates that participants report greater proportional attention to Sound on accurate predictions concerning whether a given utterance was treated as serious versus non-serious, as compared to their overall predictions. Taking a slightly different view of the data, we reach a similar conclusion by observing the accuracy rates broken out by the category of reasons given for participants' responses, i.e., by considering the rate of accuracy for each response tag independent of the proportion of the categories relative to each other. The response tags Environment, Transcript, Other, and Unsure / No Response have a relatively low number of instances in the data. As such, they are useful when viewed in aggregate in the proportional analysis above. However, when broken down by individual response tag for each excerpt, the low N tends to skew the results. Therefore, analysis of accuracy rate by response type will be limited to the five largest categories (i.e., Body Language, Content, Context, Pragmatics, and Sound). As I demonstrate, participants have a higher rate of accuracy compared

to the overall accuracy rate when they indicate that they have attended to the category Sound in their responses.

As noted above, for Survey B1, wherein they only had access to a transcript, participants chiefly stated that they had attended to Content and/or Context in responding to the survey items. As can be seen in Figure 4.38. below, these two categories track neatly with both the overall and the reasons averages.

Figure 4.38.



For the purposes of this analysis, the purple bars indicate the Overall Average of the accuracy rates for each excerpt, as seen in Figure 4.18. above, as well as all excerpts combined (two rightmost columns). The pink bars indicate the Reasons Average, which is the average accuracy rate of the reasons stated relative to the bar graph itself, for each excerpt as well as for all excerpts combined. The differing values for the Overall versus Reasons averages is attributable to the large difference in number of responses for each category tag. For example, if the response tags were theoretically evenly distributed across all response types, then the pink and purple bars would be equal. As is, because the categories Content and Context comprise such a large share of all responses, the Overall Average tends to track very closely with the average

accuracy rate of responses in these two categories, whereas the accuracy rates associated with the other categories tend to have less impact on the overall accuracy rate for each excerpt. If we wish to compare the accuracy rates of the categories to each other, the Reasons Average offers a standardized average accuracy rate, i.e., averaged across all categories, which is comparable across all response tags, regardless of variations in N.

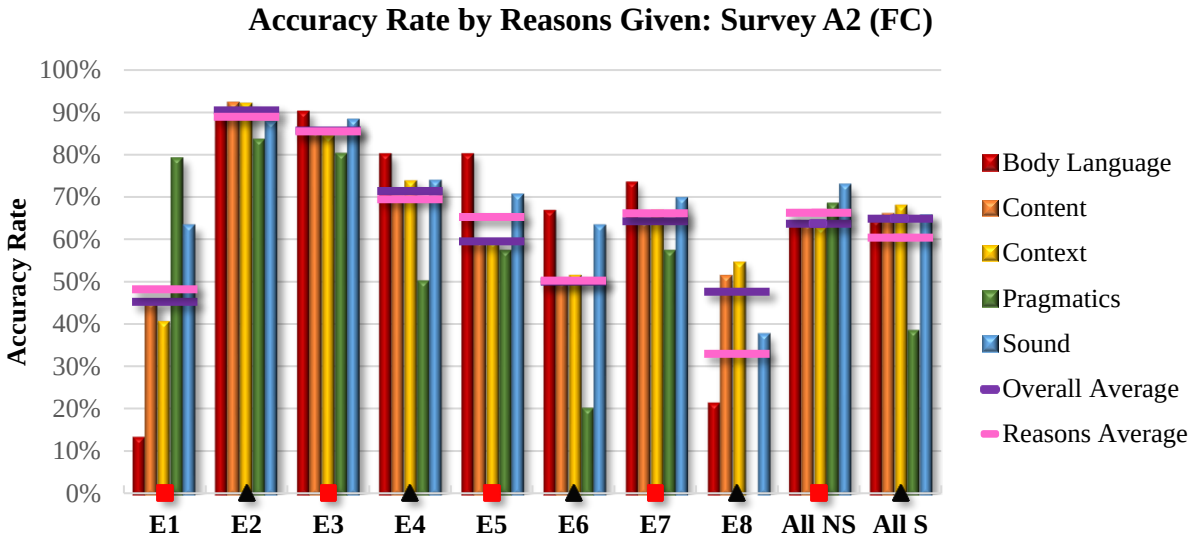
For example, in Figure 4.38. above, we note that the purple Overall Average and pink Reasons Average bars track closely together for all excerpts, with the exception of Excerpt 7. On Excerpt 7 (a non-serious utterance), the overall accuracy rate, as illustrated by the purple Overall Average bar, rests at 26% accuracy in predicting whether the utterance was treated as serious versus non-serious, which is the same accuracy rate for Excerpt 7 on Survey B1 shown in Figure 4.18. above. Despite this, the accuracy rate associated with the response tag Pragmatics is over 65%, indicating that participants who attended to pragmatic considerations in predicting whether the utterance would be treated as serious or non-serious had substantially greater accuracy in doing so as compared to participants who relied more on Content and/or Context alone. In turn, the pink Reasons Average bar jumps to just over 40%.

Although most striking in the case of Excerpt 7, the accuracy rate associated with the Pragmatics tag is above both the Overall and Reasons averages for five of the eight excerpts on Survey B1, including all four utterances treated as non-serious (Excerpts 1, 3, 5, and 7). Taken together, participants on Survey B1 tended to have higher accuracy rates in perception of non-seriousness when considering Pragmatics. This is best illustrated in the All Non-Serious column, where the green Pragmatics bar is above both average bars, and the pink Reasons Average, at about 51%, surpasses the Overall Average at around 46%. Based on this evidence, we conclude that attention to pragmatic considerations seems to play a role in perception of non-seriousness in

the transcript only survey condition, which largely confirms the conclusions reached based on the data shown Figures 4.30. and 4.31. above.

Turning now to Surveys A2 and B2, where participants had access to the video data, participants continued to rely mainly on Content and Context in their survey responses. However, consideration of the accuracy rates broken out by category tags reveals that attention to sound seems to play a larger role in successful perception of non-seriousness as compared to other categories, as seen for Survey A2 in Figure 4.39. below.

Figure 4.39.



As in Figure 4.38. above, the purple Overall Average bars indicate the overall accuracy rates for all excerpts, as found in Figure 4.18., while the pink Reasons Average bars show the average accuracy rates for all the categories relative to the bar graph. In the case of Survey A2, the purple Overall Average bars again track well with the accuracy rates of the two largest categories (Content and Context), while the pink Reasons Average bars reveal how the average accuracy rates compare across the categories. Although on Survey B1, the category Pragmatics tended to have an accuracy rate above both these averages, this is not true on Survey A2, where Pragmatics has a higher accuracy rate only on Excerpt 1. Although the category is slightly above the

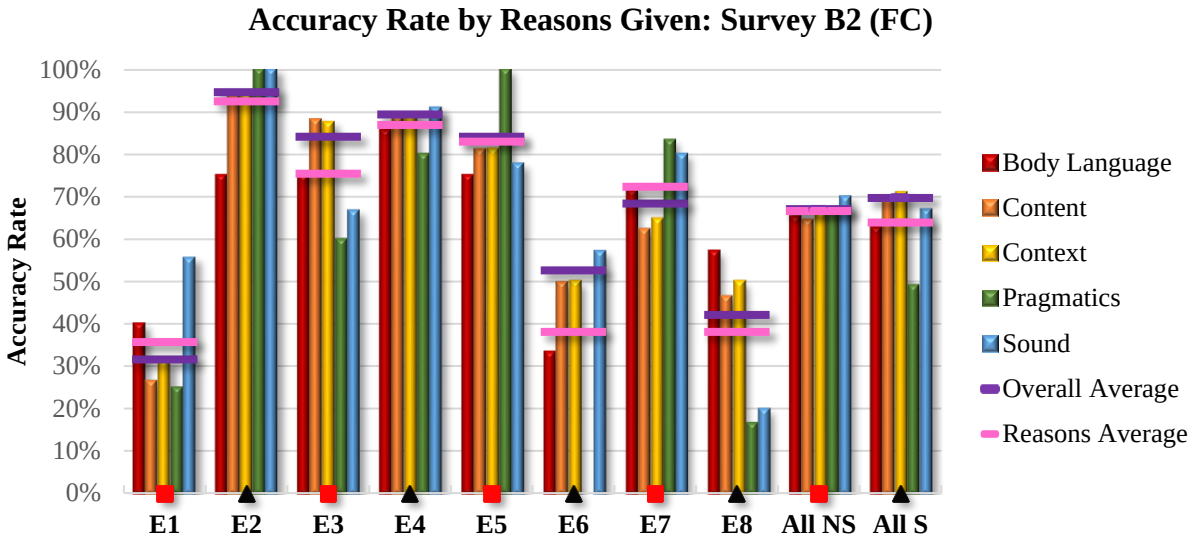
average accuracy rates for all non-serious excerpts combined, this seems largely due to the especially high accuracy rate associated with this tag on Excerpt 1, and it falls well below the average accuracy rates for all serious excerpts combined.

The accuracy rates associated with the category Body Language also are above the Overall and Reasons averages for five of the eight excerpts (Excerpts 3, 4, 5, 6, and 7). Excerpts 1 and 8 have especially low accuracy rates associated with Body Language, however, with the result that the category for all excerpts is slightly below the Overall and Reasons averages (as in the two rightmost columns). It seems that attention to Body language, then, was perhaps a useful strategy, but not successful in all instances on this survey.

The most interesting results in Figure 4.39., however, are the accuracy rates associated with the Sound tag, which are above the Overall and Reasons averages for six of eight excerpts, including all four of the non-serious excerpts. This same result holds for the Sound tag on all non-serious excerpts, with the accuracy rate associated with the category falling above the average accuracy rates. Of note is the spike in accuracy on Excerpt 1 associated with the Sound tag. In the case of the serious excerpts, the accuracy rate for Sound averages 65.5%, while the Overall and Reasons averages fall at 64.8% and 60.4%, respectively. In other words, the Sound tag is associated with higher accuracy rates on both non-serious and serious excerpts. However, the latter category is only slightly above the averages due to the low accuracy rate associated with Sound on Excerpt 8.

The results for Survey B2, in Figure 4.40. below, reveal generally similar trends to those on Survey A2.

Figure 4.40.



On Survey B2, the purple Overall Average bars again track well with the Content and Context categories, while the pink Reasons Average bars allow comparison of the accuracy rates across the categories. The Pragmatics tag still lags behind the results from Survey B1, although it is associated with slightly increased accuracy in perception of non-seriousness as compared to the results on Survey A2, with accuracy rates associated with the tag falling above the Overall and Reasons averages on three out of eight excerpts. On the other hand, the category Body Language is associated with reduced accuracy as compared to Survey A2, with accuracy rates associated with the Body Language tag falling above the Overall and Reasons averages on only two excerpts. As on Survey A2, attention to these two categories seems to be associated with improved accuracy in perception of non-seriousness with some excerpts, but not across the board.

The accuracy rates associated with the category Sound are of particular interest on Survey B2. Under this survey condition, participants had higher accuracy rates on five of the eight excerpts when they claimed to be attending to Sound, with the accuracy rate associated with this tag falling above the Overall and Reasons averages for Excerpts 1, 2, 4, 6, and 7. More

compelling still is that the accuracy rate associated with the Sound tag for all non-serious utterances also is above the Overall and Reasons averages. This lends additional support to the hypothesis that attention to Sound tends to improve accuracy in perception of non-seriousness.

Finally, of particular note are the results from both Surveys A2 and B2 for Excerpt 1, which had an especially low accuracy rate overall as compared to the other excerpts (33% on all survey conditions combined; cf. Figure 4.17.). Yet on both Surveys A2 and B2, when participants claimed to be attending to Sound, the accuracy rate on this excerpt jumps to 63% and 55.5%, respectively. In contrast, the overall accuracy rate for Excerpt 8 was 49% on all survey conditions combined (cf. Figure 4.17.). Yet among participants who claimed to be attending to Sound, the accuracy rate on this excerpt drops to 37.5% on Survey A2, and to 20% on Survey B2.

Recalling briefly the conclusion of Chapter 3, we noted that the comparison of Marion's non-serious and serious utterances was an outlier on two acoustic measures (namely, pitch and speech rate). It was suggested that survey participants might struggle to differentiate between her serious and non-serious excerpts. As noted above in Section 4.4.4., it was therefore somewhat surprising that survey participants seemed to have more success overall identifying Marion's non-serious utterance as compared to Simone's.

Yet if we consider the reasons why participants claim to have responded to the survey items as they did, it becomes clear that participants who were attending to Sound as a feature of Simone's speech were better able to correctly identify the utterance treated as non-serious. Conversely, when participants attended to Sound as a key feature of Marion's speech, they were less likely to correctly identify the utterance treated as serious. Bearing in mind the conclusions of the preceding section, wherein it was noted that although survey participants had lower

accuracy rates with Simone's Excerpt 1 as compared to Marion's Excerpt 7, in terms of signal detection, they only showed signal confusion with Marion's utterances. In other words, precisely Marion's serious utterance which was an outlier, in that it was quite similar acoustically to other participants' *non*-serious utterances, was confusing to the survey participants when they claimed to be attending to how it sounded.

This result was somewhat obscured in Section 4.4.4., due to the outsized impact of Content and Context on the overall accuracy rates. However, it becomes clear when we consider survey participants' claims about which features of the interaction were most salient to them that Excerpt 8, as an acoustic outlier, was indeed confusing for those participants who claimed to be attending to Sound. By the same token, those participants for whom the acoustic features of Excerpt 1 were particularly salient had a substantially higher rate of success on this excerpt. This is a striking result which, in conjunction with the signal detection values in Section 4.4.4., offers strong support for the claim that acoustic features play an important role in both production and perception of non-seriousness.

In summary, the preceding analysis offers two important results. First, the category Sound comprises a larger proportional share of accurate responses to the survey items. This indicates that survey participants were able to differentiate between utterances treated as serious versus non-serious at a higher rate of accuracy when they claimed to attend to the acoustic features of the utterances. Second, on Surveys A2 and B2, the Sound category has a rate of accuracy that is higher than the Overall average for most excerpts, although this effect was greater for Survey A2 than Survey B2.

Recall that for Survey B1, we noted that Pragmatics seemed to correlate with higher rates of accuracy, in a fashion similar to the role of Sound for Surveys A2 and B2. I suggest that,

although the correlation between increased attention to Sound and increased accuracy in perception of non-seriousness holds for both Surveys A2 and B2, the slight variation in degree of correlation is potentially attributable to the priming effects of Surveys A1 and B1. In other words, Survey A participants who had access to the video excerpts from the beginning might have been more inclined to continue to attend to Sound, whereas Survey B participants who had access only to transcripts initially might have continued to rely more on Pragmatics. However, as is illustrated in Figures 4.33., 4.35., and 4.37., the latter strategy (relying on Pragmatics) ultimately seems to have been less successful than the former. Taken together, these two results are clear evidence that manipulation of various acoustic features in speech has a key role in production and perception of non-seriousness.

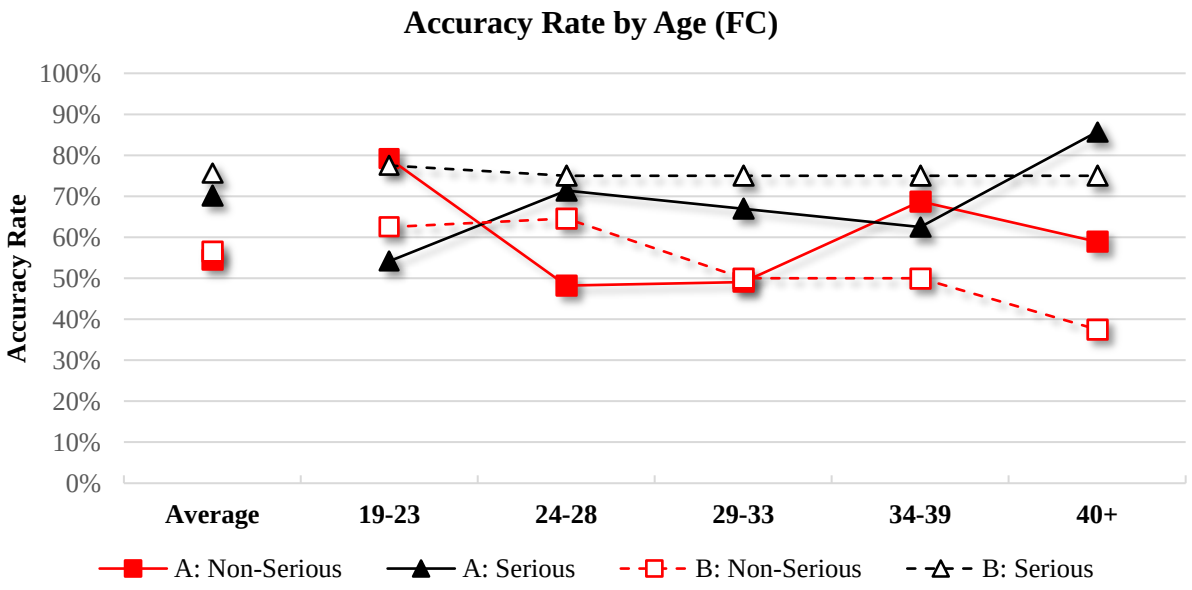
4.4.6. Survey Results by Participants' Age

To this point in the analysis, we have focused on the impact on the results of variables internal to the survey itself, i.e., which experimental group participants were assigned to, the two survey conditions particular to each group, the speakers of each excerpt, and the reasons stated by survey participants for the answers they gave. We turn now to a consideration of any potential impact of demographic variables on the results of the survey. As stated earlier in the chapter, a wealth of demographic data were collected for the study. The demographics portion of the survey can be viewed in Appendix 2.5. The demographic variables included in that survey were a selection of the most commonly investigated categories in sociolinguistic research, which will enable potential comparison of results to similar work in the field. Additional analysis of the data utilizing more sophisticated multivariate tools is required in order to determine the full impact of demographic variables on the study. However, for the purposes of demonstrating the validity of the methodology proposed in Chapter 2 for the present study, we consider only three

of the variables collected, namely: age, gender, and occupation. As argued below, the survey results seem to hold steady with little variation across these basic demographic categories.

Figure 4.41., below, shows the overall accuracy rates by age of survey participants in both experimental groups.

Figure 4.41.



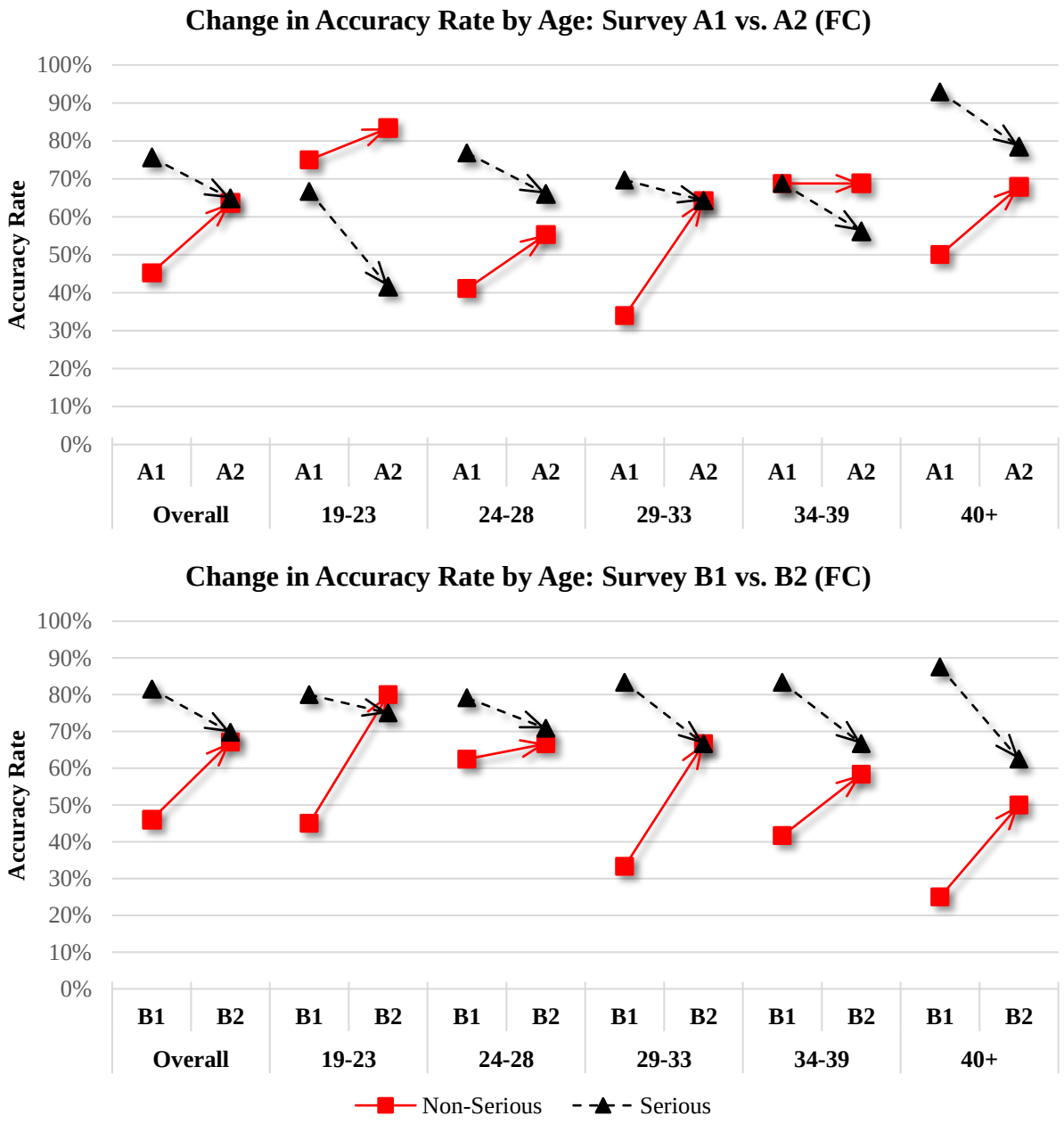
Rather than presupposing that the data would track with more traditional age groupings (i.e., 20-25, etc.), the nontraditional age groupings (i.e., 19-23, etc.) were selected based on initial analysis of the data. Specifically, these age groupings tracked more closely with the natural trends in the data, whereas the traditional intervals tended to skew the results for some age groups. These groupings therefore offered a truer approximation of how participants' responses varied by age.

The overall accuracy rates in Figure 4.41. contain the results for both surveys in each group. The results presented here track well with the results seen in Figure 4.11. in Section 4.4.2. Participants in most age groups were better able to identify serious excerpts as such in both experimental groups, with the accuracy rates for these excerpts generally higher than for the

non-serious excerpts. Comparing the results for each age group to the overall average on the far right, note that the most deviant results are in the 19-23 and 34-39 age groups, especially in the Group A experimental condition.

This observation holds when comparing the changes in results from Survey 1 to Survey 2, as seen in Figure 4.42., below.

Figure 4.42.



In Figure 4.42., the overall trends are similar to those seen previously in Figure 4.13. in Section 4.4.3. (reproduced for convenience in the far left section of the figure in the "Overall" column). Specifically, we find a convergence of accuracy rates from Survey 1 to Survey 2 towards a level accuracy rate on both utterance types across most age groups.

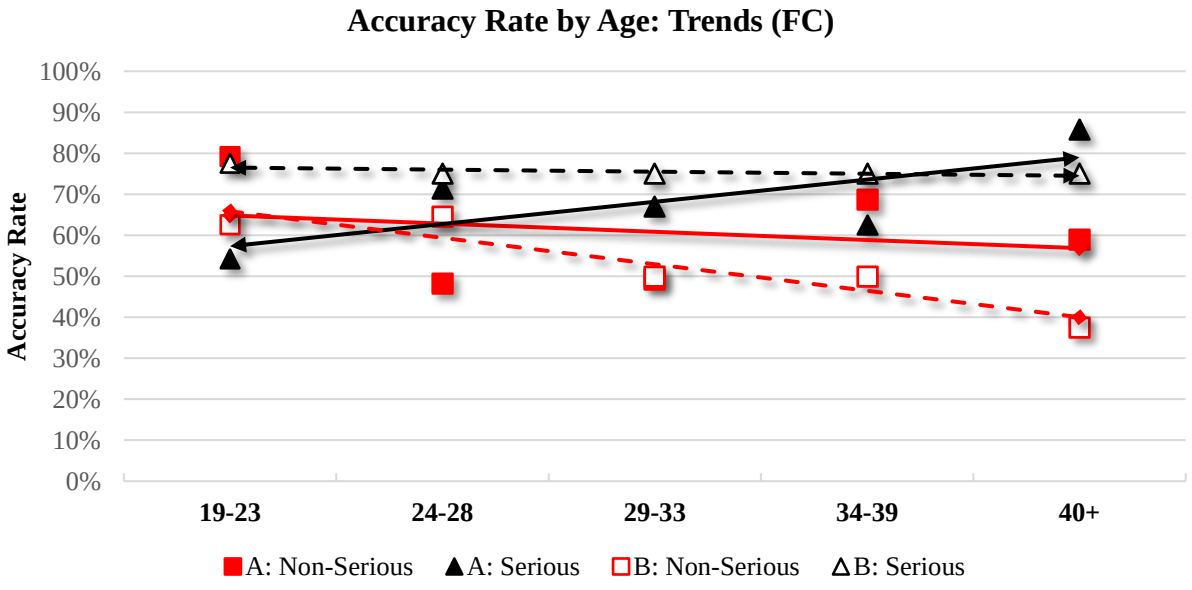
As noted above, the results which deviate most from this pattern can be seen in participants aged 19-23 and 34-39 in Group A. Here we find that participants were better able to identify non-seriousness, which is in contrast to the other participants who tend to identify serious utterances as such with better accuracy. In addition, the leveling of accuracy rates does not hold, particularly in the 19-23 age group. Prior to analysis, the running hypothesis was that age of survey participant would have little impact on the results, if any at all. If there were any variation by age, it was hypothesized that younger participants might have higher overall accuracy rates because they are in similar age groups as the speakers themselves. Although initial inspection suggests that the youngest age group might be better able to identify non-seriousness than the older groups, we must bear in mind that the task entails differentiating serious from non-serious utterances, that is, differentiating between signal and noise. Given that younger survey participants also have lower accuracy rates on the serious utterances, it does not seem to be the case that these participants are better at differentiating between serious and non-serious utterances than older participants.

In Section 4.4.3., I argued that the survey results corroborated the evidence for serious utterances being the default, unmarked condition as presented in Chapter 3. The results in Figure 4.42. above offer additional evidence in support of this argument; across all age groups, we find higher accuracy rates for serious utterances on the first survey set with a tendency towards converging accuracy rates as participants proceed to the second survey. As argued previously,

this offers additional evidence that participants in the perceptual survey tend to presume that a given utterance is serious, making [- humor] the unmarked condition. After hearing the utterances multiple times and with the additional support of transcripts, they are able to differentiate between utterances treated as serious and non-serious with greater accuracy, in effect overcoming the bias to presume that the unmarked condition is in use.

However, returning to the divergent results found in the younger cohort of Group A where we do not observe this convergence of accuracy rates, it is useful to consider the overall trends in the data as well. Figure 4.43. below reproduces the same overall accuracy rates seen previously in Figure 4.41. with average trend lines.

Figure 4.43.



The trend lines in Figure 4.43. reveal that in both experimental groups, the younger participants tend to have higher accuracy on the non-serious excerpts and lower accuracy on the serious excerpts as compared to the older participants. This phenomenon is more pronounced in the youngest cohort of Group A, which is where we initially observed the deviant changes in accuracy rates on Survey 1 and 2 in Figure 4.42., yet the observation holds for both experimental

groups that there is a tendency for higher rates of accuracy on non-serious utterances amongst the younger survey participants.

Although an interesting observation that younger participants seem more likely to notice when the marked [+ humor] condition is in use, it would be premature to conclude on this basis alone that younger participants are treating non-seriousness as the default, unmarked condition (or that older participants are any more likely to presume the reverse). The bulk of the evidence still suggests that most participants, as argued earlier in this chapter, are treating [- humor] as the default, unmarked condition, regardless of age. In short, it is difficult to draw any conclusions beyond that it is necessary to pursue further analysis using more sophisticated multivariate tools to tease out whether these are artifacts of the data (due to, for example, a smaller sample size for the 19-23 and 34-39 cohorts in Group A), or if the observed variations in accuracy with non-serious excerpts are in fact attributable to participants' age (e.g., younger survey participants were closer in age cohort to the speakers).

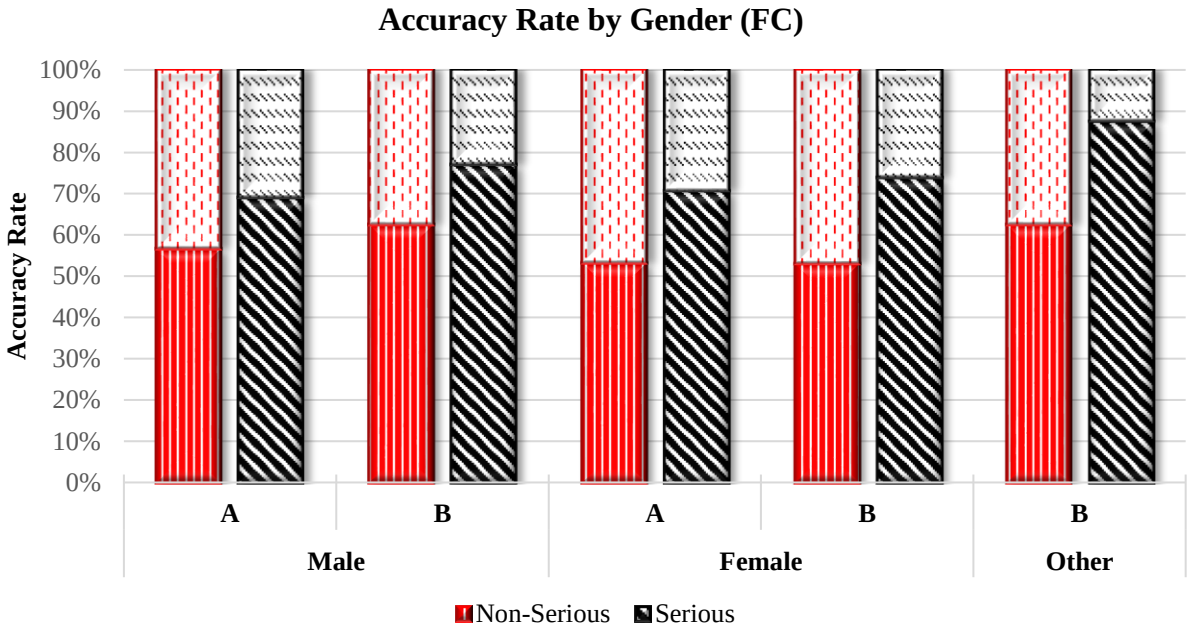
Overall we can conclude that the trends observed to this point generally hold across age groups. Participants tended to have higher accuracy rates with the serious excerpts overall, and the characteristic convergence of accuracy rates from Survey 1 to Survey 2 on both excerpt types holds as well. This offers additional evidence that participants treat [- humor] as the default, unmarked condition. Although younger participants seemed to have higher accuracy with the non-serious excerpts than older participants, especially in Group A, further analysis is required to determine if this is an artifact of the data set or attributable to participant age. At this stage, however, there is not strong evidence that responses on the perceptual survey were substantially impacted by participants' age.

4.4.7. Survey Results by Participants' Gender

In addition to age, data on survey participants' gender were also collected. Participants were asked to indicate the gender with which they identified, with options provided for male (*männlich*), female (*weiblich*), and neither (*weder noch*); the full text of this question and the corresponding answer options can be found in Appendix 2.5. As with all other demographic data collected, no attempt was made to collect information beyond what participants stated. As such, these data reflect participants' gender identity, which may or may not correspond to biological sex.

Similar to the results from the previous section, when divided out by gender, the overall trends presented in the earlier portions of this chapter generally hold. Although further analysis is required to verify the statistical power and validity of this conclusion, on the basis of this analysis, the impact of survey participants' self-identified gender appears negligible. Initial evidence of this can be viewed below in Figure 4.44., which contains the overall accuracy rates divided out by gender (cf. Figure 4.11. in Section 4.4.2. for the overall accuracy rates).

Figure 4.44.

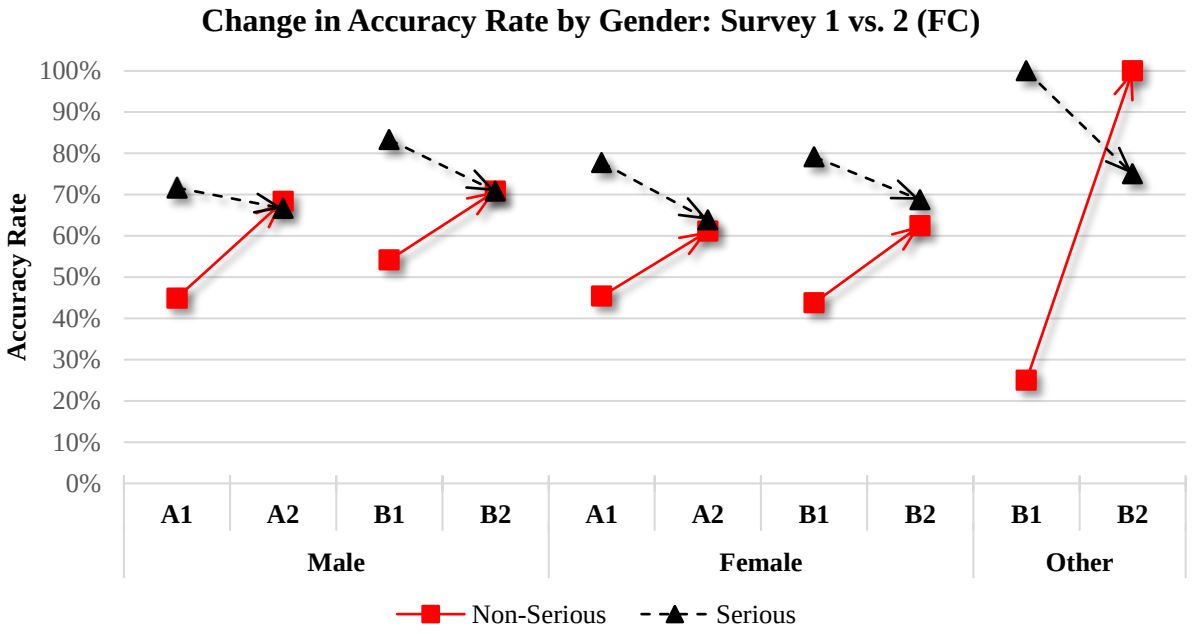


In Figure 4.44., we find that participants in both experimental groups had higher accuracy rates in identifying excerpts treated as serious, with accuracy rates hovering around 70-75%. In comparison, accuracy was lower on the excerpts treated as non-serious, with accuracy rates in the range of 55-65%. Compared to the overall accuracy rates in Figure 4.11. the results are very similar for both experimental groups.

The accuracy rates for Group B participants who identified either as male or neither gender are notably higher than the overall rates. However, this is likely attributable to the smaller number of participants who identified themselves as such. There were nearly twice as many female participants as male participants in both experimental groups, and there was only one participant who chose not to indicate gender who completed the entire survey set in full. Given that the number of Group B participants was already very small in comparison to Group A, as discussed previously in Section 4.4.3., the impact of a smaller number of male and nongendered participants is especially magnified in the Group B results.

However, regardless of the smaller number of male and nongendered participants, the general trends in the data still hold. That is, participants tended to identify excerpts treated as serious at a higher rate of accuracy in both experimental groups. Similarly, we find that these same patterns hold when considering the changes in accuracy rates over the course of the survey sets, as illustrated in Figure 4.45. below.

Figure 4.45.



The changes in accuracy rate from Survey 1 to Survey 2 follow similar patterns as found in Figure 4.13. in Section 4.4.3. Participants improved their accuracy rates on the non-serious utterances on Survey 2 while performing slightly worse on the serious utterances, such that the convergence of accuracy rates on utterances treated as serious versus non-serious over the course of Survey 1 to Survey 2 is clearly present. Despite the extremely low number of participants who did not identify with one of the presented gender options, this pattern is still visible in that group. Additional analysis would be required to verify that the slight differences in accuracy rates are attributable to gender, to the disproportionate number of participants of each gender, or

some combination thereof. However, the observed pattern of convergence in accuracy rates lends support to the analysis that the slightly higher accuracy rates seen in Figure 4.44. for the male and nongendered participants in Group B are more likely a result of the smaller sample sizes as opposed to representing any sort of statistically significant variation within those groups.

As with the analysis by age cohort in Section 4.4.6., there do not at this point appear to be any major fluctuations in the data on the basis of gender identity. It was not anticipated that there would be any great impact on results due to age or gender identity, which the data have borne out. In aggregate, these results support the overall conclusion that participants, regardless of demographic cohort, are treating seriousness as the default condition.

4.4.8. Survey Results by Participants' Profession

The final demographic variable considered in this analysis is profession. When designing the survey, it was anticipated that professional linguistic training might impact the results, especially for linguists working in CA and related fields who might be better able to detect subtle cues in conversation which would indicate that an utterance was likely to be treated by participants as non-serious. As such, a question about participants' professional field of study and/or employment was included in the demographics portion of the survey.

For the purposes of analysis, participants were assigned to the categories of linguist or non-linguist based on the text entered in response to the questions about major field of study and current and/or future career (see Appendix 2.5. for the full text of the demographics survey). If a participant entered either a major area of study or a career which indicated training in any subfield of linguistics, then they were assigned to the category "linguist". A broad view of linguistic training was taken; for example, participants who majored or minored in linguistics, even if they did not continue with a career in linguistics, were included in the linguist category,

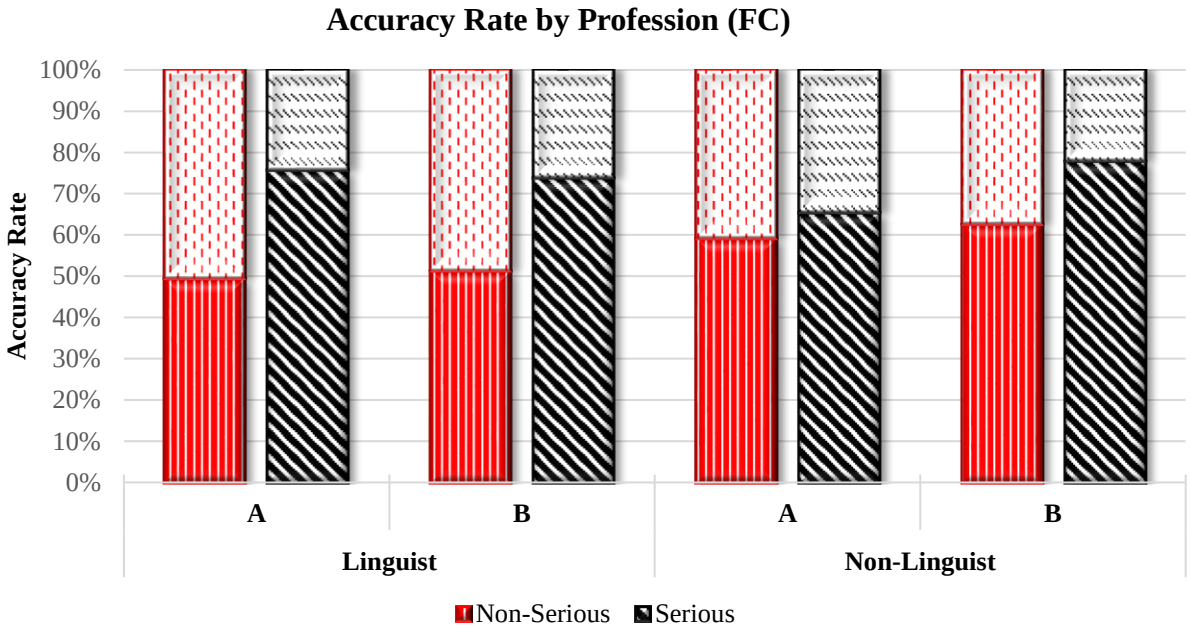
as were participants who did not major or minor in linguistics, but who had experience with practical applications of linguistics, such as teaching language (e.g., German as a second language).

Once survey participants had been coded as either linguists or non-linguists using the coding scheme described above, it was found that nearly half of all participants (approximately 49%) could be considered linguists. This is obviously a much higher proportion of the participant pool than would be found in the general population. Based on a preliminary scan of the linguist population, there are pockets of linguists who are of similar age and are majoring or minoring in linguistics at a handful of German universities. It would appear then that linguistic instructors at these universities shared the survey link with their students, who then in turn comprised a rather large portion of the survey participants.

With such a relatively large sample set of linguist participants (as compared to the general population), their results are in fact more likely to be representative of the linguist demographic group because the larger sample set offers a less skewed snapshot of the population. By the same token, the low number of non-linguist participants relative to general population, in particular, warrants further analysis in future to determine what impact this smaller participant pool might have had on the accuracy rates seen in Figures 4.46. and 4.47., below. At this time, it was not feasible to run a fuller analysis which differentiates the impact on the results of various subfields of linguistics (such as CA, as noted above), however, the data are available for future analysis.

The results of the analysis comparing survey results for linguists versus non-linguists are illustrated in Figure 4.46., below.

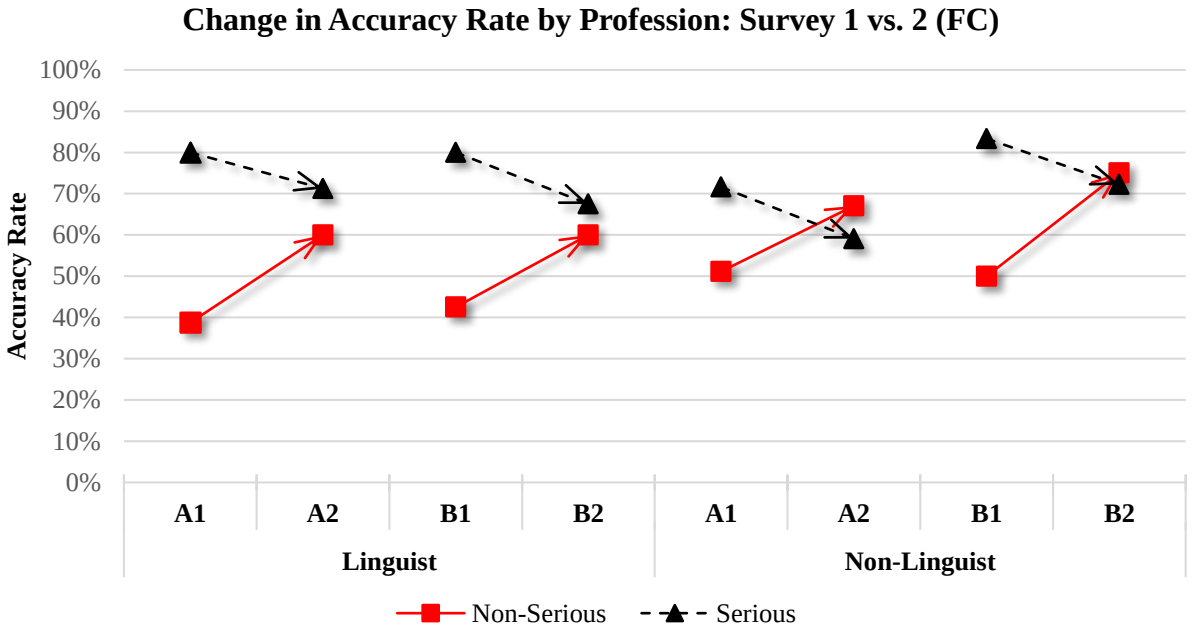
Figure 4.46.



First note that non-linguists have higher accuracy rates on the non-serious excerpts as compared to linguists in both experimental groups, a finding which is contrary to the anticipated results described above, i.e., that linguists might have higher accuracy. Comparing these results to those found in Figure 4.11. in Section 4.4.2., we find that the linguists are slightly below the averages for both Group A and Group B on the non-serious excerpts (approximately 55% for both groups), but above average on the serious excerpts (approximately 70% and 75%, respectively). In contrast, the non-linguists are above the averages for both Group A and Group B on the non-serious excerpts, but are below or close to the averages on the serious excerpts.

In addition to the overall accuracy rates being relatively in line with earlier results, the changes in accuracy rates from Survey 1 to Survey 2 are also similar, as seen in Figure 4.47. below.

Figure 4.47.



Comparing these results to Figure 4.13. in Section 4.4.3., the same pattern of convergence of accuracy rates from Survey 1 to Survey 2 on excerpts treated as non-serious versus those treated as serious is seen here. Note that the linguist group again has slightly lower accuracy rates on non-serious excerpts as compared to the non-linguist group, while performing slightly better on the serious excerpts. Despite the differences in raw accuracy rates, the same patterns of convergence are present in both groups, so we can conclude that both demographic groups performed similarly on the survey set in both experimental groups A and B, that is, on the second survey they tended to have higher accuracy with the excerpts treated as non-serious and lower accuracy with the excerpts treated as serious.

In terms of signal detection, linguists and non-linguists in both experimental groups displayed greater signal sensitivity on the second survey as compared to the first. The linguists show a larger improvement in signal detection over the course of the survey set in Group A (A' = 0.680 (A1:Ling) and 0.740 (A2:Ling)) as compared to the non-linguists (A' = 0.690 (A1:Non-

Ling) and 0.708 (A2:Non-Ling)). On the other hand, the non-linguists showed greater improvement in Group B (A' = 0.767 (B1:Non-Ling) and 0.821 (B2:Non-Ling)) than the linguists (A' = 0.703 (B1:Ling) and 0.716 (B2:Ling)). In sum, the linguists performed slightly better than the non-linguists in terms of signal detection in Group A (A' = 0.709 (A:Ling) and 0.697 (A:Non-Ling)), while the non-linguists performed better in Group B (A' = 0.707 (B:Ling) and 0.791 (B:Non-Ling)). Considering the response bias of both groups, we find that the linguists tended to display a negative response bias, that is, they tended to assume that the utterances were treated as serious (B'' = 0.195 (A1:Ling), 0.079 (A2:Ling), 0.209 (B1:Ling), and 0.045 (B2:Ling)), whereas the non-linguists were more likely to assume that the utterances were treated as non-serious (B'' = 0.103 (A1:Non-Ling), -0.045 (A2: Non-Ling), 0.286 (B1: Non-Ling), and -0.034 (B2: Non-Ling)).

As noted above, it had been anticipated that linguists would be better able to predict when utterances were treated as serious versus non-serious than non-linguists. However, this was not borne out in the data: non-linguists had slightly better accuracy rates overall. For both experimental groups, the general movement from Survey 1 to Survey 2 was similar to the overall trends observed through the rest of the chapter, for both linguists and non-linguists. Similarly, there were not large differences in signal sensitivity, with linguists performing slightly better than non-linguists in Group A, and vice versa in Group B. In short, it appears that training in linguistics did not to have a great impact effect on the survey results. This was a surprising, yet pleasing, result, which shows that, if anything, the measurement tool was effective.

4.5. Conclusion

This chapter opened with the argument that, from a CA perspective, accurate perception of the [+/- humor] status of a given utterance is vital for conversationalists determining how to

respond. Specifically, I noted that conversational contiguity following a non-serious utterance is maintained solely by the deployment of laughter subsequent to the utterance; any other response has the potential to be treated as a source of trouble (Jefferson, 1979; Sacks, 1972). Conversely, laughter subsequent to a [- humor] utterance, with no serious uptake of the action of the original utterance, may also be treated as a source of trouble (Sacks, 1972). Given the importance of participants' accurate perception of [+/- humor] status, I further argued that recipient design could be deployed at the phonetic level, whereby speakers may mark an utterance phonetically as non-serious, rather than relying on content and/or context alone. This view dovetails nicely with a more recent research trend in sociolinguistics which emphasizes the importance of verifying the results of any production analysis by testing whether the linguistic features under study are also perceptible and salient for everyday users of the language. Such perceptual experiments serve as a control which decreases the impact of the analyst's own perceptual biases on the results of the production experiment, while also potentially undergirding the results of the same (Thomas, 2002).

In terms of research on humor, there have been some studies which have tested perception of non-seriousness. The vast majority of these have utilized data collected in a laboratory. These data frequently consisted of nonsense syllables, or were digitally manipulated to remove some phonetic features, rendering the utterances incomprehensible in terms of content, while preserving the general pitch contours. In short, there has been a dearth of perceptual experiments using naturally occurring conversational data to test perception of non-seriousness. The sole exception to this, as noted above, were the Bryant and Fox Tree (2002, 2005) studies. Despite this reliance on linguistically impoverished data, in aggregate, these experiments offered tantalizing evidence that phonetic cues may play some role in perception of non-seriousness, at

least in the context of a laboratory.

The results of the perceptual experiment presented in this chapter sought to close this gap by testing the perceptibility and salience of non-seriousness in naturally occurring data. Based on the results presented in this chapter, it is clear that untrained listeners can identify at a rate greater than chance would predict whether participants in the original conversation are likely to treat a given utterance as non-serious by laughing in response. These results held across both experimental groups under both survey conditions for all data excerpts, and did not vary based on participants' age, gender, or profession.

The most striking results stemmed from analysis of the reasons participants gave for their responses to the survey items. Here, we found that participants had higher accuracy in differentiating between utterances treated in the original conversations as serious versus as non-serious when they claimed to be attending to phonetic cues, as collected under the response category Sound. This result held when the responses were taken together for the survey as a whole, and also when broken down for each excerpt individually: participants in the perceptual survey were better able to identify non-seriousness when they stated that they had considered the sound of the utterance in responding to the survey items.

This result was especially enlightening in the case of Marion's serious utterance (Excerpt 8). As described in Chapter 3, Marion's utterances were acoustic outliers in terms of changes in pitch and speech rate when comparing her serious and non-serious excerpts. As seen in Figures 4.39. and 4.40., participants who claimed to be attending to sound on Excerpt 8 had substantially lower accuracy rates. This in turn lends particularly strong support to the hypotheses that phonetic cues likely play a vital role in marking utterances as [+/- humor], and that such marking is indeed perceptible and salient for listeners. I return to this point in an integrated analysis of

the results of both the production and perceptual experiments in Chapter 5.

The hypotheses that survey participants could, absent full conversational context, differentiate between utterances treated as serious versus non-serious by participants in naturally occurring conversation, and that they would be able to do so with greater accuracy when presented with acoustic data, appear to be supported by these data. Although statistical significance was not measured due to the small data set, the trends in the data are suggestive that further study is warranted. On the basis of these results, perceptual experiments utilizing naturally occurring data can play a key role in advancing our understanding of non-seriousness, in particular experiments with a larger data set in order to determine statistical significance. The strong indications for the perceptibility and salience of acoustic marking of non-seriousness should be given special attention in any future research.

Chapter 5

The Problem of Non-Seriousness: What Next?

> "Well he isn't the type of person who would understand anything like that, you know? I've seen other people - as a bartender he meets other people like this, and I've seen him at work and people will try to tell him something and he'll listen and then he'll make a big joke of it, and things like this, you know? He doesn't mean any harm by it, but-" (Sacks, 1972, p. 71)

5.1. Introduction

For any given utterance in any given conversation, we must, almost instantaneously, reach a decision about how to react. Often, we do so by laughing, and in doing so, we treat the prior utterance as non-serious (Sacks, 1972). Although certainly we can, and do, misjudge whether an utterance is best treated as serious or non-serious, for example, by laughing at inappropriate times (Jefferson, 2010) or by failing to laugh when it would be appropriate (Haddington, 2011; Jefferson, 2010), by and large, the vast majority of sequences containing laughter are not treated as sources of trouble. This offers support to the view that we are, generally speaking, quite adept at identifying non-seriousness when it occurs, such that we can respond appropriately, that is, by laughing.

Despite the evidence that we overwhelmingly know when to laugh (and vice versa, when *not* to laugh), there is no clear consensus as to *how* we know this. Although much is known about the rhetorical and sequential features of non-seriousness, it is not always the case that

content and context alone are sufficient information to permit accurate determination of the [+/-humor] status of an utterance. I first began collecting German conversational data by telephone in 2007, and later video data in Germany in 2008-09. In those datasets, I observed that many of the utterances treated as non-serious seemed to be made with a voice which differed appreciably from surrounding utterances treated as serious from the same speaker(s). I hypothesized that speakers may potentially be manipulating a variety of features of the acoustic stream in order to mark an utterance as non-serious. Much research has been done on the acoustic features of, for example, irony (e.g., Anolli, Ciceri, & Infantino, 2000; Attardo & Pickering, 2011), sarcasm (e.g., Voyer & Techentin, 2010), laughter (e.g., Bachorowski, Smoski, & Owren, 2001; Kipper & Todt, 2003; LaGreca, Laforest, Alioua, & Vincent, 1996; Vettin & Todt, 2004), and "smile voice" (e.g., Auberge & Cathiard, 2003; Tartter, 1980; Tartter & Braun, 1994). Although most of these studies concluded that, when making humorous utterances, speakers do manipulate the acoustic stream in ways which differentiate it from surrounding baseline talk, the literature has ultimately offered conflicting results in regard to which acoustic features are manipulated, and how.

I set out to address this gap and to discover how precisely it is that we, as speakers of our respective languages, know when to laugh during naturally occurring everyday conversation. Guided by the best practices available in Conversation Analysis (CA) and sociophonetics, I devised a methodology utilizing data culled from naturally conversations in German (Chapter 2). The results indicate that there are indeed measurable acoustic differences between utterances treated as non-serious as compared to surrounding utterances treated as serious (Chapter 3), and that these differences are perceptible and salient to untrained listeners (Chapter 4). Yet I believe that the greater contribution this work has to offer is the methodology itself. With broad

applicability, this methodology could yield results which are more generalizable and comparable across studies, not only for research on the production and perception of humor, but for virtually any topic of interest in sociolinguistics.

5.2. A Methodology for the Analysis of Non-Seriousness

In Chapter 2, I argued that neither CA nor sociophonetics in isolation are best suited to the study of non-seriousness. Traditionally, we have relied on an explanation of humor grounded in its rhetorical, philosophical, and/or psychological features and functions (e.g., Attardo, 1994). According to this view, humor is broadly understood to be accomplished via an unexpected rupture from the prior statement, that is, the juxtaposition of the content of an utterance with its context (e.g., Attardo & Raskin, 1991; Dynel, 2009). (Socio)linguists have tested these theories using various means, such as breaching experiments (Bell, 2009) and composing scripts to test whether they are understood as ironic (e.g., Gerrig & Goldvarg, 2000; Kreuz & Link, 2002; Kreuz & Roberts, 1995). Others have studied the acoustic markers of humorous utterances (e.g., Anolli et al., 2000; Attardo & Pickering, 2011), and whether these are perceptible to listeners as indicators, for example, of irony and/or sarcasm (e.g., Voyer & Techentin, 2010) or of smiling while speaking (e.g., Auberge & Cathiard, 2003; Tartter, 1980; Tartter & Braun, 1994).

Although these studies provide evidence of acoustic markers of non-seriousness, the vast majority of this research relies on laboratory data, or on data collected from scripted television and radio programs. Given the wide variation in data sources used in these sociophonetic studies, it is unsurprising that the results have been conflicting, if promising. As we know from CA, if we want to understand how humans interact in everyday conversation, we should make everyday conversation the object of study by using it as our source of data (Sacks, 1995a, 1995b). However, in the case of humor and jokes, CA researchers again largely focused on the

content and context of humor by documenting the sequential features of non-seriousness (e.g., Norrick, 1993; Sacks, 1972; Schegloff, 2001) and laughter (e.g., Jefferson, 1972, 1979, 1984, 1985, 2004, 2010; Jefferson, Sacks, & Schegloff, 1987). This sequential focus has revealed much about how non-seriousness is organized on a turn-by-turn basis in naturally occurring conversation, and how it is used at the interpersonal level. However, the question of how we identify non-seriousness as such remains open, especially in cases where the content and context of an utterance are insufficient means to determine whether it is serious or non-serious.

Building on and integrating the best practices of CA and sociophonetics to address this question, I proposed a combined methodology for this study. This tactic was directly inspired by the work of Couper-Kuhlen and Selting (1996), Golato and Fagyal (2008), Kaimaki (2011), and Wennerstrom (2001), who have all called for a similarly integrated approach with increased precision in measuring the acoustic and phonetic features of conversational data, while maintaining the data collection procedures and sequential focus of more traditional CA work. Building on their approach, and inspired by the work of Kerswill and Wright (1990) and Thomas (2002), I have introduced additional experimental controls in line with current trends in (socio)phonetic and acoustic analysis (e.g., Jannedy & Hay, 2006).

In summary, this proposed methodology relies on implementation of CA-style data collection procedures for obtaining naturally occurring conversational data, as well as sequential analyses in order to amass a collection of utterances which can be used as experimental variables and controls. These utterances are then subjected to acoustic analysis in order to locate differences between the two sets. Finally, the same utterances are incorporated into a perceptual survey in order to ascertain whether these differences are perceptible and salient to untrained listeners, and not just a result of the analyst's own perceptual biases. As argued in Chapter 2,

this procedure satisfies the methodological strictures of both CA and sociophonetics, while also allowing us to investigate the interplay between phonetic production and sequential actions with more readily verifiable and generalizable results.

5.3. Production of Non-Seriousness

For the production branch of this study, the working hypothesis was as follows:

H_{1A}: Utterances which are identifiable as non-serious on the basis of how they are treated by participants in a naturally-occurring conversation are produced with acoustic features which differ measurably from surrounding serious talk from the same speaker.

In order to test this hypothesis, a set of excerpts treated as non-serious were selected using CA-style sequential analyses. Once selected, a corresponding set of utterances treated as serious were located in the immediately preceding talk from the same speakers. The collection of utterances were then tested in PRAAT for the following acoustic parameters: pitch, pitch variation, vowel formants, intensity, beats, and speech rate. The results for the non-serious and serious utterances were compared to locate any differences between the two sets of utterances for each speaker, and to determine if there were any similarities in how the utterances differed across speakers.

The results indicated that three out of four speakers had decreased average pitch, decreased range of pitch, and decreased variation in pitch in their non-serious utterances as compared to the serious ones. Overall, the average pitch values and range seemed to collapse towards a narrower, as-yet-undefined norm. Furthermore, the speech rate decreased in three of four speakers. In the case of pitch and speech rate, the outlier was Marion's utterances (Excerpts 7/8), who had increases in pitch average, range, and variation, and in speech rate. On the other hand, the results for vowel formants, intensity, and beats all split their results, tracking with the

regional varieties of German spoken by the participants. The speakers from Berlin-Brandenburg had backer vowels, increased intensity average, range, and variation, and decreased beats per second, whilst the speakers from Hessen had the opposite results on all of these parameters (fronter vowels, decreased intensity average, range, and variation, and increased beats per second).

Overall, there are observable trends in how speakers manipulate the acoustic stream in order to mark an utterance as non-serious, and as such, we can reject the null hypothesis. In particular, the results for pitch and speech rate seem especially promising, with three out of four speakers following similar patterns. However, it is worth underscoring again that the results for vowel formants, intensity, and beats per second all consistently tracked with the regional varieties of German spoken by the participants.

5.4. Perception of Non-Seriousness

The working hypotheses for the perception branch of this study were as follows:

H_{1B}: Absent full conversational context, participants in a survey can identify utterances which were treated as non-serious in the original conversation at a rate greater than chance would predict.

H_{1C}: When presented with acoustic data, participants in a survey can identify utterances which were treated as non-serious in the original conversation at a greater rate of accuracy than when presented with transcripts only.

Participants in the perceptual survey viewed the same excerpts measured for the acoustic analyses, although the excerpts were trimmed prior to the reactions of the original conversation participants' to the utterances under study. They were then asked to indicate what would happen next in the conversation, and what had just happened, by selecting from a series of randomized options. On the second half of the survey, participants were prompted to indicate why they had

responded as they did.

Under both survey conditions, Group A (audio-visual only to audio-visual with transcript condition) and Group B (transcript only to audio-visual with transcript condition) both appear to predict whether or not the utterances had been treated as non-serious by the original conversation participants. In addition, both groups had greater accuracy rates on the second half of the survey with the non-serious utterances, with slightly reduced accuracy with serious utterances. These overall results suggest that untrained listeners can indeed predict which utterances were originally treated as non-serious. This effect was generally consistent across various demographic groups, including age cohort, gender, and profession.

The most promising results came from analysis of the perceptual results broken out for each speaker, and of the reasons survey participants' gave for their responses. Participants had the highest accuracy and signal detection rates with speakers Maximilian (Excerpts 3/4) and Gerhart (Excerpts 5/6). With speaker Simone (Excerpts 1/2), participants had lower accuracy rates, but high signal detection, indicating that they could differentiate between the serious and non-serious utterances, despite some confusion about the non-serious utterance itself. For speaker Marion (Excerpts 7/8), the opposite was true: participants had somewhat higher accuracy rates, but greater signal confusion, indicating that they could not differentiate between her serious and non-serious utterances. The analysis of the reasons participants gave for their responses revealed a correlation between higher accuracy rates in differentiating between speakers' serious and non-serious utterances and survey participants' claims of attending to the sound of the excerpt. However, this finding was especially striking in the case of Marion (Excerpts 7/8), where participants had substantially lower accuracy on her serious excerpt when they stated that they had considered the sound of her utterance.

In aggregate, these results suggest that, absent full conversational context, participants can differentiate between utterances treated as serious versus non-serious. There is also evidence to suggest that they are able to do so with greater accuracy when they have access to the acoustic data, in that the survey participants had greater accuracy rates when they claimed to have considered the sound of the utterance in giving their responses to the survey questions. The generalizability of these results is limited because the sample set was too small to allow measurement of statistical significance. However, these preliminary results lend support to the original hypotheses, and suggest that further study is warranted.

5.5. Conclusions

Pulling together the results from the production and perception branches of this study, as summarized above, the key findings suggest that there are likely acoustic markers of non-seriousness which are perceptible and salient to untrained listeners. I first review the most promising findings in a unified analysis that integrates the results of both the production and perception branches of this study, as inspired by Schaffer (1981), and then suggest some future avenues of research to pursue, both in terms of potential expansions on this study, and of applications of the findings to other, related areas of inquiry.

As noted, the most promising acoustic results were found with the changes in pitch average, range, and variation, and speech rate, which were all decreased in three out of four speakers. The outlier was Marion, who had increased pitch average, range, and variation, and speech in her non-serious utterance (Excerpt 7) as compared to her serious one (Excerpt 8). On the perception survey, participants had middling to high accuracy rates in identifying her non-serious utterance, but due to lower accuracy rates on her serious utterance, there was a high level of signal confusion with her excerpts. Furthermore, participants had lower accuracy on Excerpt

8 (Marion's serious utterance) when they claimed to be attending to sound. Comparing the acoustic measurements to the results from the perception experiment, it becomes clear that the serious excerpt seems to have greater similarity to the non-serious excerpts from the other speakers, in that it has lower pitch average, range, variation, and speech rate, and is fairly consistently heard as non-serious by the survey participants. In other words, the results taken together offer strong support that untrained listeners may perceive changes in pitch and speech rate as a cue for non-seriousness, in particular when those changes are decreases.

Additionally, as has been well-documented in the CA literature, dispreferred utterances are often made more slowly, often with hesitations and gaps (e.g., Pomerantz, 1984; Sacks, Schegloff, & Jefferson, 1974). Although traditionally this has been understood as applying to dispreferred next actions, that is, dispreferred responses, as I argued in Chapter 1, the system of preference described in CA research bears strong similarity to the concept of markedness in linguistics more broadly. Building on Sacks (1972), I further argued in Chapter 4 that the survey results offer ample evidence to suggest that, for the category [+/- humor], the default, unmarked form is [- humor]. As such, I submit that it would not be a surprising outcome if the marked form [+ humor] would also be subject to the general observation that dispreferred (or marked) forms are issued more slowly than surrounding talk. Given that the results from the survey suggest that decreased speech rate correlates strongly with participants' perception of an utterance as [+ humor], this dovetails nicely with the already well-established finding from CA that dispreferred utterances are generally made more slowly than the surrounding talk.

Further evidence in support of the view that speakers manipulate the acoustic stream to mark an utterance as non-serious is found in the results for vowel formants. As described above, the results indicated that the speakers from Berlin-Brandenburg had backer vowels, while the

speakers from Hessen had fronter vowels. That these results track well with variety of German spoken is already promising, however, additional evidence to support this view again comes from the survey results. Survey participants had the highest rates of accuracy and signal detection with the utterances from speakers Maximilian (Excerpts 3/4) and Gerhart (Excerpts 5/6). On the other hand, they had lower accuracy rates and good signal detection with the utterances from speaker Simone (Excerpts 1/2), and middling accuracy rates and poor signal detection with the utterances from speaker Marion (Excerpts 7/8). Comparing the vowel formants results for all speakers, as detailed in Chapter 3, Maximilian (Excerpts 3/4) and Gerhart (Excerpts 5/6) had quite substantial shifts in the vowel formants of their non-serious utterances as compared to their serious ones, while Simone (Excerpts 1/2) and Marion (Excerpts 7/8) had more subtle shifts, especially in the /i/ vowel. Taken together, there is an apparent correlation of higher rates of accuracy and signal detection of non-seriousness with larger changes in vowel formants, which offers additional evidence in support of the argument that such changes in vowel formants are potentially perceptible and salient for untrained listeners.

The final acoustic parameters measured were intensity and beats. As described in Chapter 3, there was again a clear correlation in changes in intensity and beats with regional variety of German spoken. Specifically, the speakers from Berlin-Brandenburg had increased intensity average, range, and variation, and decreased beats per second in the non-serious excerpts as compared to their serious ones, while speakers from Hessen had decreased intensity average, range, and variation, and increased beats per second in the non-serious excerpts. This finding is interesting due to the regional variation already seen with vowel formants. However, the changes in intensity and beats were fairly consistent in degree across all speakers, whereas the changes in vowel formants were greater (and therefore more likely perceptible and salient for

untrained listeners) in the utterances from the speakers with higher accuracy and signal detection rates in the perceptual survey (i.e., speakers Maximilian (Excerpts 3/4) and Gerhart (Excerpts 5/6)). As such, it is not possible to conclude at this time without further statistical analysis whether the changes in intensity and beats are regionally-determined markers of non-seriousness which are perceptible and salient for untrained listeners, or are simply artifacts.

5.6. Limitations and Future Directions

As just noted, a limitation of this research is the sample size in both the production and perception branches of the study. Due to the small data set, the results are limited to general observations about trends in the data, and it is not possible to run analyses to determine whether the results are statistically significant, which limits the generalizability of the results. A further limitation is the survey design itself, which was potentially taxing for participants and may have inadvertently reduced the sample size as a result. Finding ways to streamline the survey could greatly increase participation and reduce attrition, which would in turn increase the sample size.

In addition, careful attention must also be given to the broader action sequences of excerpts selected for use in any similar work. For example, although all of the excerpts selected for use here were clear examples of utterances treated as non-serious, two of the utterances occurred in a sequential context of hypothetical discourse (Excerpts 1/2 and 7/8), while the other two occurred in a sequential context of complaining (Excerpts 3/4 and 5/6). Furthermore, use of telephone data in both branches of the study may have allowed for better control of variables with regard to the potential role of body language and facial expressions in production and perception of non-seriousness. Finally, the distractors used on the survey may require reworking in order to determine whether they are truly plausible responses to the questions. Specifically, a better design would be to select only those distractors which are potentially fitted to the context

of each excerpt, rather than providing the same set of distractors for all excerpts.

Turning now to future directions for the investigation of non-seriousness, as argued here, the most promising findings regarding acoustic cues of non-seriousness in German conversation are pitch and speech rate. Three out of four speakers manipulated these acoustic features in similar ways, and those results correlated well with the perceptual results. Similarly, greater changes in vowel formants also correlated with higher rates of accuracy and signal detection on the perceptual survey, supporting the conclusion that changes in vowel formants may serve as a regionally-determined cue for non-seriousness. The least promising results were found with changes in intensity and beats, where there was again regional variation in changes in these acoustic features, but there was not clear evidence from the perceptual survey that these changes were salient for untrained listeners.

It is possible that changes in any of these acoustic features could be done in either direction to signal non-seriousness. A larger dataset is required to determine whether these changes are consistently more successful in one direction than the other, or if changes in these features in any direction is sufficient to serve as an acoustic marker of non-seriousness. For example, in this set, decreases in both pitch and speech rate appear to have been more salient for survey participants, and were therefore a more successful strategy for marking an utterance as non-serious. It is possible that Marion (Excerpts 7/8) was also attempting to manipulate her pitch and speech rate, but that increased pitch and speech rate were a less viable strategy. In other words, it may be the case that *any* changes in pitch and speech rate can be used as cues for non-seriousness, but that changes in one direction or the other, or perhaps collapsing towards a normative pitch range, as observed here and in the pilot study, are simply more successful strategies.

In light of these findings, the most important next step is to run the same acoustic analyses on the corresponding lab data from each speaker, specifically on the scripted sentences the participants read aloud in serious and non-serious modes (as described in Chapter 2; cf. Appendix 2.3.). Particular attention should be given to the results for pitch, speech rate, and vowel formants, although certainly additional insight might still be gleaned from consideration of intensity and beats. This step would allow analysis of how natural versus scripted non-seriousness in German compare to each other (cf. Golato, 2003). In addition, it is vital to run similar sequential and acoustic analyses on a larger set of excerpts from more speakers, and to run additional perceptual surveys for verification of the results. In parallel, the results should be compared to similar work using telephone data in order to determine what role embodiments, such as body language and facial expressions, play in production and perception of non-seriousness.

Although a survey of the sort designed for this study could prove too taxing for participants if it contained more excerpts, it may be possible to design a less taxing survey which nonetheless includes more excerpts than were used here. For example, the survey design implemented by Bryant and Fox Tree (2002) is one way in which a survey that has more excerpts but is less taxing could be designed. In that survey, participants only listened to the utterance in question, and then read the original context and a composed context. The composed context was designed to be opposite of the original context in terms of serious versus ironic. In other words, originally serious utterances were presented both with the original serious context, and with a composed ironic context, while original ironic utterances were presented both with the original ironic context, and a composed serious context. Then, participants were asked to identify which context seemed most appropriate for the utterance they heard. Their results

indicated that participants were able to identify the original context, whether serious or ironic, based on the sound of the original utterance alone. This survey design may be substantially less taxing, and could therefore offer a means of scaling up the verification, via perceptual survey, of a larger study of the acoustic features of non-seriousness. However, caution should be exercised in selecting any survey design requiring substantial amounts of reading, especially given the results in Chapter 4, where it was found that assignment to the Group B experimental condition with transcripts correlated with higher attrition rates.

With a larger data set, it would be possible to determine, for example, whether the changes in acoustic features described in Chapter 3 are seen on a larger scale, especially with regard to any regional variation. Furthermore, a larger data set would allow for more sophisticated statistical analyses in order to determine whether or not the changes in acoustic features and the perceptual survey results are statistically significant. In particular, a larger data set is necessary to run a multivariate analysis, which would permit consideration of whether survey participants who are from the same regions in Germany as are the speakers in the excerpts are better able to differentiate between utterances originally treated as serious versus non-serious. Such an analysis would test the hypothesis that the acoustic cues of non-seriousness may vary by region and/or dialect, which emerged from the results presented in Chapter 3.

More broadly, this study should be replicated using naturally occurring data in other languages, which would greatly enrich our understanding of the acoustic features of non-seriousness (or humor) as these compare cross-linguistically. These results offer valuable insight into the production and perception of non-seriousness in German conversation. However, if other languages mark non-seriousness acoustically, it may be that this is done in entirely differ ways than can be found in German language conversation.

Finally, as argued at the outset of this chapter, the greatest contribution of this study is the methodological framework itself. The delicate interplay between sequential and phonetic features of naturally occurring conversation remains, at this time, relatively understudied, with sociophonetics generally relying on laboratory data, and conversation analysts generally relying on impressionistic transcription methods which neglect to account for the analysts' perceptual biases and filters. This study design addresses this methodological gap, and could be adapted to investigate a broad range of topics that are of interest to sociolinguists of any stripe. As I expand this research to a larger data set and continue to refine the methodology, I hope that fellow linguists will find unique, creative, and fruitful applications for this methodological proposal.

Bibliography

Anolli, L., Ciceri, R., & Infantino, M.G. (2000). Irony as a game of implicitness: Acoustic profiles of ironic communication. *Journal of Psycholinguistic Research, 29*(3), 275-311.

Archakis, A., Giakoumelou, M., Papazachariou, D., & Tsakona, V. (2010). The prosodic framing of humour in conversational narratives: Evidence from Greek data. *Journal of Greek Linguistics, 10*, 187-212.

Attardo, S. (1994). *Linguistic theories of humor*. Berlin: Mouton de Gruyter.

Attardo, S. (2000). Irony markers and functions: Towards a goal-oriented theory of irony and its processing. *RASK, 12*, 3-20.

Attardo, S. (2002). Humor and irony in interaction: From mode adoption to failure of detection. In L. Anolli, R. Ciceri, & G. Riva (Eds.), *Say not to say: New perspectives on miscommunication* (pp. 159-179). Amsterdam: IOS Press.

Attardo, S. (Ed.). (2017). *The Routledge handbook of language and humor*. New York: Routledge.

Attardo, S., Attardo, D. H., Baltes, P., & Petray, M. J. (1994). The linear organization of jokes: Analysis of two thousand texts. *Humor, 7*(1), 27-54.

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor, 16*(2), 243-260.

Attardo, S., & Pickering, L. (2011). Timing in the performance of jokes. *Humor, 24*(2), 233-250.

Attardo, S., Pickering, L., & Baker, A. (2011). Prosodic and multimodal markers of humor in conversation. *Pragmatics & Cognition, 19*(2), 224-247.

Attardo, S., & Raskin, V. (1991). Script theory revis(it)ed: Joke similarity and joke representation model. *Humor, 4*(3/4), 293-347.

Auberge, V., & Cathiard, M. (2003). Can we hear the prosody of smile? *Speech Communication, 40*, 87-97.

Auer, J. C. P., & Uhmann, S. (1982). Aspekte der konversationellen Organisation von Bewertungen. *Deutsche Sprache, 1*, 1-32.

Bachorowski, J., Smoski, M. J., & Owren, M. J. (2001). The acoustic features of human laughter. *The Journal of the Acoustical Society of America*, *110*(3), 1581-1597.

Ball, P. (2014, November 18). FC Barcelona: The science behind their success. *BBC*. Retrieved from http://www.bbc.com/.

Barnes, B., Palmary, I., & Durrheim, K. (2001). The denial of racism: The role of humor, personal experience, and self-censorship. *Journal of Language and Social Psychology*, *20*(3), 321-338.

Bell, N. D. (2009). Impolite responses to failed humor. In N. R. Norrick & D. Chiaro (Eds.), *Humor in Interaction* (pp. 143-163). Amsterdam: John Benjamins.

Bertrand, R., & Priego-Valverde, B. (2011). Does prosody play a specific role in conversational humor? *Pragmatics & Cognition*, *19*(2), 333-356.

Betz, E. (2017). Diskursmarker aus konversationsanalytischer Sicht: Prosodisch integriertes *ja* am Beginn von responsiven Turns. In H. Blühdorn, A. Deppermann, H. Helmer, & T. Spranz-Fogasy (Eds.), *Diskursmarker im Deutschen: Reflexionen und Analysen* (pp. 183-205). Göttingen: Verlag für Gesprächsforschung.

Bird, C. (2011). Formulaic jokes in interaction: The prosody of riddle openings. *Pragmatics & Cognition*, *19*(2), 268-290.

Bryant, G. A. (2004). Prosodic features of verbal irony in spontaneous speech (Doctoral dissertation, University of California Santa Cruz). Available from ProQuest Dissertation and Thesis database. (UMI No. 3135046).

Bryant, G. A. (2010). Prosodic contrasts in ironic speech. *Discourse Processes*, *47*(7), 545-566.

Bryant, G. A. (2011). Verbal irony in the wild. *Pragmatics & Cognition*, *19*(2), 291-309.

Bryant, G. A., & Fox Tree, J. E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol*, *17*(2), 99-117.

Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech*, *48*(3), 257-277.

Cheang, H. S. (2008). The sounds of sarcasm in English and Cantonese: A cross-linguistic production and perception study (Doctoral dissertation, McGill University). Available from ProQuest Dissertation and Thesis database. (UMI No. NR50795).

Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication, 50*, 366-381.

Cheang, H. S., & Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. *Journal of the Acoustical Society of America, 126*(3), 1394-1405.

Cheang, H. S., & Pell, M. D. (2011). Recognizing sarcasm without language: A cross linguistic study of English and Cantonese. *Pragmatics & Cognition*, *19*(2), 203-223.

Chiaro, D. (2009). Cultural divide or unifying factor? Humorous talk in the interaction of bilingual, cross-cultural couples. In N. R. Norrick & D. Chiaro (Eds.), *Humor in Interaction* (pp. 211-231). Amsterdam: John Benjamins.

Chomsky, N. (1965). *Aspects of the theory of syntax*. The Hague: Mouton.

Comey, J. B., Jr. (2018, April 25). CNN Town Hall: Comey: Truth, Lies and Leadership. Interview by A. H. Cooper. In *CNN Live Event Special* [Television broadcast]. Video retrieved April 26, 2018, from CNN website, https://www.cnn.com/videos/politics/2018/04/26/james-comey-town-hall-full.cnn.

Couper-Kuhlen, E. (1993). *English speech rhythm: Form and function in everyday verbal interaction.* Amsterdam: John Benjamins.

Couper-Kuhlen, E. (2004). Analyzing language in interaction: The practice of *never mind*. *English Language and Linguistics*, *8*(2), 207-237.

Couper-Kuhlen, E. (2009). Relatedness and timing in talk-in-interaction. In D. Barth-Weingarten, N. Dehé, & A. Wichmann (Eds.), *Where prosody meets pragmatics* (pp. 257-276). Bingley: Emerald Group Publishing.

Couper-Kuhlen, E., & Selting, M. (1996). Towards an interactional perspective on prosody and a prosodic perspective on interaction. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation: Interactional studies* (pp. 11-56). Cambridge: Cambridge University Press.

Cummins, F. (n.d.). *Fred's Miscellaneous Collection: Instructions for estimating the locations of beats in a soundfile* [sic]. Retrieved September 4, 2014, from University College Dublin website, http://cspeech.ucd.ie/Fred/beatExtraction.php.

Drew, P. (1987). Po-Faced receipts of teases. *Linguistics, 25*, 219-253.

Dynel, M. (2009). *Humorous garden-paths: A pragmatic-cognitive study*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Ervin-Tripp, S. M., & Lampert, M. (2009). The occasioning of self-disclosure humor. In N. R. Norrick & D. Chiaro (Eds.), *Humor in Interaction* (pp. 3-27). Amsterdam: John Benjamins.

Flamson, T., Bryant, G. A., & Barrett, H. C. (2011). Prosody in spontaneous humor: Evidence for encryption. *Pragmatics & Cognition*, *19*(2), 248-267.

Ford, C. E., & Fox, B. A. (2010). Multiple practices for constructing laughables. In D. Barth-Weingarten, E. Reber, & M. Selting (Eds.), *Prosody in interaction* (pp. 339-368). Amsterdam: John Benjamins.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs: Prentice-Hall.

Gerhardt, C. (2009). Multimodal and intertextual humor in the media reception situation: The case of watching football on TV. In N. R. Norrick & D. Chiaro (Eds.), *Humor in Interaction* (pp. 79-98). Amsterdam: John Benjamins.

Gerrig, R. J., & Goldvarg, Y. (2000). Additive effects in the perception of sarcasm: Situational disparity and echoic mention. *Metaphor and Symbol*, *15*(4), 197-208.

Glenn, P. (1989). Initiating shared laughter in multi-party conversations. *Western Journal of Speech Communication*, *53*, 127-149.

Glenn, P. (1991). Current speaker initiation of two-party shared laughter. *Research on Language and Social Interaction*, *25*, 139-162.

Glenn, P., & Holt, E. (Eds.). (2013). *Studies of laughter in interaction*. London: Bloomsbury.

Glenn, P., & Holt, E. (2017). Conversation Analysis of Humor. In S. Attardo (Ed.), *The Routledge handbook of language and humor* (pp. 295-308). New York: Routledge.

Goffman, E. (1959). *The presentation of self in everyday life*. New York: Anchor Books.

Goffman, E. (1971). *Relations in public: Microstudies of the public order*. New York: Basic Books.

Golato, A. (2003). Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics*, *24*(1), 90-121.

Golato, A. (2012). Impersonal quotation and hypothetical discourse. In K. Heylen & N. Stukker (Eds.), *Converging evidence in language and communication research* (pp. 3-36). Amsterdam: John Benjamins.

Golato, A. (2018). Turn-initial *naja* in German. In J. Heritage & M. L. Sorjonen (Eds.), *Between turn and sequence: Turn-Initial particles across languages* (pp. 413-444). Amsterdam: John Benjamins.

Golato, A., & Fagyal, Z. (2008). Comparing single and double sayings of the German response token *ja* and the role of prosody: A Conversation Analytic perspective. *Research on Language and Social Interaction*, *41*(3), 241-270.

Goldstone, J. (Producer), & Jones, T. (Director). (1983). *Monty Python's the meaning of life* [Motion picture]. United Kingdom: Universal Studios.

Graham, E. E., Papa, M. J., & Brooks, G. P. (1992). Functions of humor in conversation: Conceptualization and measurement. *Western Journal of Communication*, *56*, 161-183.

Haakana, M. (2002). Laughter in medical interaction: From quantification to analysis, and back. *Journal of Sociolinguistics*, *6*(2), 207-235.

Haddington, P. (2011). Serious or non-serious? Sequential ambiguity and disavowing a prior stance. *Functions of Language*, *18*(2), 149-182.

Haugh, M. (2010). Jocular mockery, (dis)affiliation, and face. *Journal of Pragmatics*, *42* (8), 2106-2119.

Haugh, M., & Bousfield, D. (2012). Mock impoliteness, jocular mockery and jocular abuse in Australian and British English. *Journal of Pragmatics*, *44*(9), 1099-1114.

Haverkate, H. (1990). A speech act analysis of irony. *Journal of Pragmatics*, *14*(1), 77-109.

Hay, J. (2001). The pragmatics of humor support. *Humor*, *14*(1), 55-82.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458-484.

Hepburn, A., & Varney, S. (2013). Beyond ((laughter)): Some notes on transcription. In P. Glenn & E. Holt (Eds.), *Studies of laughter in interaction* (pp. 25-38). London: Bloomsbury.

Heritage, J. (1990). Intention, meaning, and strategy: Observations on contraints on interaction analysis. *Research on Language and Social Interaction*, *24*, 311-332.

Heritage, J. (1998). Oh-Prefaced responses to inquiry. *Language in Society*, *27*, 291-334.

Heritage, J., & Atkinson, J. M. (1984). Introduction. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in Conversation Analysis* (pp. 1-15). Cambridge: Cambridge University Press.

Hidalgo Navarro, A. (2011). Humor, prosodia e intensificación pragmática en la conversación colloquial española. *Verba, 38*, 271-292.

Hirst, D., & Di Cristo, A. (1998). *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press.

Holmes, J. (2000). Politeness, power and provocation: How humour functions in the workplace. *Discourse Studies, 2*(2), 159-185.

Holmes, J., & Marra, M. (2002). Humor as a discursive boundary marker in social interaction. In A. Duszak (Ed.), *Us and others: Social identities across languages, discourses and cultures* (pp. 377-400). Amsterdam: John Benjamins.

Holt, E. (2013a). Conversation Analysis and Laughter. In C. A. Chapelle, *The encyclopedia of applied linguistics* (pp. 1-6). Malden: Blackwell Publishing.

Holt, E. (2013b). "There's many a true word said in jest": Seriousness and nonseriousness in interaction. In P. Glenn & E. Holt (Eds.), *Studies of laughter in interaction* (pp. 69-90). London: Bloomsbury.

Jannedy, S., & Hay, J. (2006). Modelling sociophonetic variation. *Journal of Phonetics, 34*(4), 405-408.

Jefferson, G. (1972). Side sequences. In D. Sudnow (Ed.), *Studies in social interaction* (pp. 294-338). New York: The Free Press.

Jefferson, G. (1979). A technique for inviting laughter and its subsequent acceptance declination. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 79-96). New York: Irvington Publishers.

Jefferson, G. (1984). On the organization of laughter in talk about troubles. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in Conversation Analysis* (pp. 346-359). Cambridge: Cambridge University Press.

Jefferson, G. (2004). A note on laughter in 'male-female' interaction. *Discourse Studies, 6*(1), 117-133.

Jefferson, G. (2010). Sometimes a frog in your throat is just a frog in your throat: Gutturals as (sometimes) laughter-implicative. *Journal of Pragmatics, 42*(6), 1476-1484.

Jefferson, G., Sacks, H., & Schegloff, E. (1987). Notes on laughter the pursuit of intimacy. In G. Button & J. R. E. Lee (Eds.), *Talk and social organisation* (pp. 152-205). Clevedon: Multilingual Matters.

Kaimaki, M. (2011). Transition relevance and the phonetic design of English call openings. *Journal of Pragmatics*, *43*(8), 2130-2147.

Kerswill, P., & Wright, S. (1990). The validity of phonetic transcription: Limitations of a sociolinguistic research tool. *Language Variation and Change*, *2*(3), 255-275.

Kersten, K. (2009). Humor and interlanguage in a bilingual elementary school setting. In N. R. Norrick & D. Chiaro (Eds.), *Humor in Interaction* (pp. 187-210). Amsterdam: John Benjamins.

Kipper, S., & Todt, D. (2003). The role of rhythm and pitch in the evaluation of human laughter. *Journal of Nonverbal Behavior*, *27*(4), 255-272.

Kohler, K. J. (2008). 'Speech-smile', 'speech-laugh', 'laughter' and their sequencing in dialogic interaction. *Phonetica*, *65*(1-2), 1-18.

Kotthoff, H. (2009). An interactional approach to irony development. In N. R. Norrick & D. Chiaro (Eds.), *Humor in Interaction* (pp. 49-77). Amsterdam: John Benjamins.

Kreuz, R. J. (1996). The use of verbal irony: Cues and constraints. In J. S. Mio & A. N. Katz (Eds.), *Metaphor: Implications and applications* (pp. 23-38). Mahwah: Lawrence Erlbaum.

Kreuz, R. J., & Link, K. E. (2002). Asymmetries in the use of verbal irony. *Journal of Language and Social Psychology*, *21*(2), 127-143.

Kreuz, R. J., & Roberts, R.M. (1995). Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Metaphor and Symbolic Activity*, *10*(1), 21-31.

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*, 273-309.

Labov, W. (1966). *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh & J. Sherzer (Eds.), *Language in use* (pp. 28-53). Englewood Cliffs: Prentice-Hall.

Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of language*. Malden: Blackwell Publishing.

LaGreca, S., Laforest, M., Alioua, A., & Vincent, D. (1996). Laughter as interaction strategy: Discursive and phonetic aspects. In J. Arnold, R. Blake, B. Davidson, S. Schwenter, & J. Solomon (Eds.), *Sociolinguistic variation: Data, theory, and analysis: Selected papers*

from NWAV 23 at Stanford (pp. 313-332). Stanford: Center for the Study of Language and Information.

Linneweber, J. (2016). Why a German 'oh' is not necessarily an English 'oh': Showing understanding and emotions with change-of-state tokens. *Die Unterrichtspraxis*, *49*(2), 186-198.

Lowe, S. (2010, July 9). Spain's "tiki-taka" style dominates. *Sports Illustrated*. Retrieved from http://www.si.com/.

MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah: Lawrence Erlbaum Associates.

Mandelbaum, J. (2012). Storytelling in conversation. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 492-507). Chichester: Wiley-Blackwell.

Markaki, V., Merlino, S., Mondada, L., & Oloff, F. (2010). Laughter in professional meetings: The organization of an emergent ethnic joke. *Journal of Pragmatics*, *42*(6), 1526-1542.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

Norrick, N. R. (1993). *Conversational joking: Humor in everyday talk*. Bloomington: Indiana University Press.

Norrick, N. R. (2001). On the conversational performance of narrative jokes: Toward an account of timing. *Humor*, *14*(3), 255-274.

Norrick, N. R. (2003). Issues in conversational joking. *Journal of Pragmatics*, *35*(9), 1333-1359.

Norrick, N. R. (2004). Humor, tellability, and conarration in conversational storytelling. *Text, 24*(1), 79-111.

Norrick, N. R. (2010). Laughter before the punch line during the performance of narrative jokes in conversation. *Text & Talk*, *30*(1), 75-95.

Norrick, N. R., & Bubel, C. (2009). Direct address as a resource for humor. In N. R. Norrick & D. Chiaro (Eds.), *Humor in Interaction* (pp. 29-47). Amsterdam: John Benjamins.

Nwokah, E. E., Hsu, H., Davies, P., & Fogel, A. (1999). The integration of laughter and speech in vocal communication: A dynamic systems perspective. *Journal of Speech, Language, and Hearing Research, 42,* 880-894.

Ochs, E. (1979). Transcription as theory. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics* (pp. 43-72). New York: Academic Press.

Ogden, R. (2013, March). *Studies from the periphery of language: Why phonetics is more than larynx-to-lips*. Paper presented at the 17th Working Conference on Research on Conversation and Discourse (17. Arbeitstagung zur Gesprächsforschung), Mannheim, Germany.

Ogden, R. (2015). Data always invite us to listen again: Arguments for mixing our methods. *Research on Language and Social Interaction, 48*(3), 271-275.

Pickering, L., Corduas, M., Eisterhold, J., Seifried, B., Eggleston, A., & Attardo, S. (2009). Prosodic markers of saliency in humorous narratives. *Discourse Processes, 46*(5), 517-540.

Pillet-Shore, D. (2012). Greeting: Displaying stance through prosodic recipient design. *Research on Language Social Interaction, 45*(4), 375-398.

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred / dispreferred turn shapes. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in Conversation Analysis* (pp. 57-101). Cambridge: Cambridge University Press.

Potter, J. (2002). Two kinds of natural. *Discourse Studies, 4*(4), 539-542.

Potter, J., & Hepburn, A. (2010). Putting aspiration into words: 'Laugh particles', managing descriptive trouble and modulating action. *Journal of Pragmatics, 42*(6), 1543-1555.

Priego-Vlaverde, B. (2009). Failed humor in conversation: A double voicing analysis. In N. R. Norrick & D. Chiaro (Eds.), *Humor in Interaction* (pp. 167-183). Amsterdam: John Benjamins.

Purandare, A., & Litman, D. (2006). Humor: Prosody analysis and automatic recognition for F*R*I*E*N*D*S*. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 208-215.

Purnell, T., & Salmons, J. (2006, March 25). Interview by M. Feldman. *Whad'Ya Know?* [Radio broadcast]. Retrieved May 7, 2012, from Michael Feldman's Whad'Ya Know? website, http://www.notmuch.com/Show/Archive.pl?s_id=394.

Repp, B. H. (1981a). Phonetic and auditory trading relations between acoustic cues in speech perception: Preliminary results. *Haskins Laboratories Status Report on Speech Research, 67-68*, 165-189.

Repp, B. H. (1981b). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Haskins Laboratories Status Report on Speech Research, 67-68,* 1-40.

Repp, B. H. (1983a). Phonetic and auditory trading relations between acoustic cues in speech perception: Further results. *Haskins Laboratories Status Report on Speech Research, 73,* 121-139.

Repp, B. H. (1983b). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication, 2*(4), 341-361.

Repp, B. H. (1983c). Trading relations among acoustic cues in speech perception: Speech-specific but not special. *Haskins Laboratories Status Report on Speech Research, 76,* 129-132.

Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research, 29*(5), 483-495.

Sacks, H. (1972). An initial investigation of the usability of conversational data for doing sociology. In D. Sudnow (Ed.), *Studies in social interaction* (pp. 31-74). New York: The Free Press.

Sacks, H. (1973). On some puns: With some intimations. *Georgetown University Monograph Series on Languages and Linguistics, 25,* 135-144.

Sacks, H. (1974). An analysis of the course of a joke's telling in conversation. In R. Bauman & J. Sherzer (Eds.), *Exploration in the ethnography of speaking* (pp. 337-353). Cambridge: Cambridge University Press.

Sacks, H. (1978). Some technical considerations of a dirty joke. In J. Shenkein (Ed.), *Studies in the organization of conversational interaction* (pp. 249-269). New York: Academic Press.

Sacks, H. (1987). On the preferences for agreement and contiguity in sequences in conversation. In G. Button & J. R. E. Lee (Eds.), *Talk and social organisation* (pp. 54-69). Clevedon: Multilingual Matters.

Sacks, H. (1995a). *Lectures on conversation: Volume I.* Malden: Blackwell Publishing.

Sacks, H. (1995b). *Lectures on conversation: Volume II.* Malden: Blackwell Publishing.

Sacks, H., Schegloff, E. A., Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696-735.

Schaffer, R. (1981). Are there consistent vocal cues for irony? *Papers from the Parasessions, Chicago Linguistic Society*, 204-210.

Schegloff, E. A. (1987). Some sources of misunderstanding in talk-in-interaction. *Linguistics, 25*, 201-218.

Schegloff, E. A. (2001). Getting serious: Joke → serious 'no'. *Journal of Pragmatics*, *33*(12), 1947-1955.

Schegloff, E. A. (n.d.). *Sound Clips for Schegloff's Publications*. Retrieved November 17, 2015, from University of California Los Angeles, Social Sciences website, http://www.sscnet.ucla.edu/soc/faculty/schegloff/sound-clips.html.

Schenkein, J. N. (1972). Towards an analysis of natural conversation and the sense of *heheh*. *Semiotica, 6*(4), 344-377.

Schmitt, R., & Knöbl, R. (2013). *Recipient design* aus multimodaler Sicht. *Deutsche Sprache, 41*(3), 242-276.

Schnurr, S., & Holmes, J. (2009). Using humor to do masculinity at work. In N. R. Norrick & D. Chiaro (Eds.), *Humor in Interaction* (pp. 101-123). Amsterdam: John Benjamins.

Schwitalla, J. (2001). Lächelndes Sprechen und Lachen als Kontextualisierungsverfahren. In K. Adamzik & H. Christen (Eds.), *Sprachkontakt, -vergleich, -variation* (pp. 325-344). Tübingen: Niemeyer.

Sherzer, J. (1985). Puns and jokes. In T. A. van Dijk (Ed.), *Handbook of discourse analysis (Vol. 3): Discourse and dialogue* (pp. 213-221). London: Academic Press.

Stafford, L., Burggraf, C. S., & Sharkey, W. F. (1987). Conversational memory: The effects of time, recall, mode, and memory expectancies on remembrances of natural conversations. *Human Communication Research, 14*(2), 203-229.

Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, *27*(1), 24-27.

Tartter, V. C., & Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America, 96*(4), 2101-2107.

Thomas, E. R. (2002). Sociophonetic applications of speech perception experiments. *American Speech¸ 77*(2), 115-147.

Trask, R. L. (1999). *Key concepts in language and linguistics*. London: Routledge.

Trouvain, J., & Truong, K. P. (2014). Towards unravelling prosodic characteristics of speaker-overlapping laughing in conversational speech corpora. In D. Barth-Weingarten & B. Szczepek Reed (Eds.), *Prosodie und Phonetik in der Interaktion* (pp. 115-135). Mannheim: Verlag für Gesprächsforschung.

Trygstad, B. J. (2014, June). *Yes, Seriously: Joke to Serious 'Ja (Yes)' in German Conversation*. Paper presented at the 4th International Conference on Conversation Analysis, Los Angeles, CA.

UEFA Euro Cup results. (2012). Retrieved July 22, 2015, from UEFA website, http://www.uefa.com/uefaeuro/season=2012/teams/team=122/matches/index.html.

Urios-Aparisi, E., & Wagner, M. (2011). Prosody of humor in *Sex and the City*. *Pragmatics & Cognition*, *19*(3), 507-529.

Veatch, T. C. (1998). A theory of humor. *Humor*, *11*(2), 161-215.

Vettin, J., & Todt, D. (2004). Laughter in conversation: Features of occurrence and acoustic structure. *Journal of Nonverbal Behavior*, *28*(2), 93-115.

Vine, B., Kell, S., Marra, M., & Holmes, J. (2009). Boundary-marking humor: Institutional gender and ethnic demarcation in the workplace. In N. R. Norrick & D. Chiaro (Eds.), *Humor in Interaction* (pp. 125-139). Amsterdam: John Benjamins.

Vogel, A. P., Maruff, P., Snyder, P. J., & Mundt, J. C. (2009). Standardization of pitch-range settings in voice acoustic analysis. *Behavior Research Methods*, *41*(2), 318-324.

Voyer, D., & Techentin, C. (2010). Subjective auditory features of sarcasm. *Metaphor and Symbol*, *25*(4), 227-242.

Walker, G. (2017). Visual representations of acoustic data: A survey and suggestions. *Research on Language and Social Interaction*, *50*(4), 363-387.

Wennerstrom, A. (2001). *The music of everyday speech: Prosody and discourse analysis*. Oxford: Oxford University Press.

Wiese, R. (1996). *The phonology of German*. New York: Clarendon Press, Oxford University Press.

Zifonun, G., Hoffmann, L., & Strecker, B. (1997). *Grammatik der deutschen Sprache*. Berlin: Walter de Gruyter.

Appendices

Appendix 2.1. Consent Forms and Debriefing Text

Appendix 2.1.1. Conversational Data Collection Phase (English translation)

Consent Form: Social Actions in German Conversation

You are invited to participate in a study that investigates social actions in conversational German as spoken in Germany. We are Joseph Salmons and Brandy Trygstad. Joseph Salmons is a professor and Brandy Trygstad is a graduate student in the Department of German at the University of Wisconsin-Madison. You were selected as a possible participant in this study because you are a speaker of Standard German and are over the age of 18. You will be one of up to 50 subjects chosen to participate in this study.

If you decide to participate in this study, you and a second person of your choosing (for example, a friend or family member) will be recorded in audio and video while you carry on an everyday conversation, for example, on the telephone, at dinner, or during social gatherings, and also while reading aloud a script in a language lab. All conversational partners must agree to participate in the study and sign a consent form. Each recording would take up to between one and three hours and we would audio- and/or video- record you three times if you decide to participate in this study. The first recording would take place in a location of your choosing, while the second and third recording would take place at the University of Potsdam (Campus Neues Palais). After the completion of all recordings, the data will be analyzed linguistically. In addition, select excerpts from the video data may be included in an electronic questionnaire.

The risks involved in this study are minimal and are comparable to risks in everyday life. Since this study investigates everyday conversations, it is possible that you might talk about personal matters with your family, and you might therefore sometimes feel uncomfortable with being audio- and video-taped in your conversations. Additionally, it is possible that participants who complete the electronic questionnaire could potentially recognize your face. There is no cost to you for participating in this project. If any abuse or neglect is witnessed or suspected while in your home, the researchers may be required to report this.

There will be no direct benefit to you for participating in this research. You will receive no payment for your participation in this study. However, this study hopes to provide you with an indirect benefit by contributing to our knowledge of how language is used and how native speakers of German interact. If you would like to receive a copy of the results of the research at no cost, you can contact Joseph Salmons or Brandy Trygstad at the phone numbers or e-mail addresses provided below.

Please note that any information that is obtained in connection with this study and that can be

identified with you will remain confidential and will be disclosed only with your permission. More specifically, all video- and audiocassettes will be coded so that no personally identifying information is on the cassette labels. Any transcriptions of the tapes will use code names and numbers, and the excerpts from the video data which will be included in the electronic questionnaire will be selected such that they do not contain any private or especially embarrassing information. Additionally, the video data will be protected against download by third parties. Thus, your confidentiality will be protected as much as possible. We will keep all cassettes in a secure place such as a locked file cabinet. All tapes recorded during this research project and photocopies of data collected and transcribed will be used for research and data analysis purposes only. The tapes will not be released to anyone, including other researchers, without your written permission (you can give your consent to this below). You have the right to review tapes and transcripts of recorded interactions and/or interviews in which you participate, and you have the right to prohibit disclosure of any portions of tapes or transcripts.

Following analysis, we will keep the tapes in a secure place for further related and unrelated research purposes conducted by us. In the future, very brief excerpts of the tapes and transcripts might be used for research publications if you give your consent to this below. All data used for these purposes will be coded to ensure the protection of your identity.

Your decision whether or not to participate will not affect your future relations with us or the University of Wisconsin-Madison. You are under no obligation to participate in this study. You are free to (a) discontinue participation in the study at any time, (b) request that the video camera or cassette recorder be turned off at any time during the recording, and (c) request that an already recorded session be destroyed and thus excluded from the study.

Your participation is completely voluntary. Your signature indicates that you have read and understood the information provided above and have decided to participate. You are free to withdraw your consent and discontinue your participation at any time after signing this form without consequences.

If you have any questions now, please ask me. If you need any additional information later, please do not hesitate to contact us. You can reach us at: 818 Van Hise Hall, 1220 Linden Drive, Madison, WI 53706, USA; telephone (608) 262-2192; E-mail: jsalmons@wisc.edu, trygstad@wisc.edu. Should you have any questions regarding your rights as a research subject, you can contact the Institutional Review Board Office, 310 Lathrop Hall, Madison WI 53706, USA; telephone (collect): (608) 263-2320, fax: (608) 890-3900; Email: lmlarson@ls.wisc.edu.

_____
Name of Participant

_____ _____
Signature of Participant Date

_____ _____
Signature of Investigator Date

Please answer the following three questions by checking off the yes/no responses and by

signing your initials. You have the right to review tapes and transcripts of recorded interactions in which you participate, and you have the right to prohibit disclosure of any portions of tapes or transcripts.

1. I grant the investigator permission to use excerpts out of the original audio and video recordings and the transcribed data **at professional meetings and in professional publications**. Any personal or sensitive information will be removed, and any name or place references will be changed.

   [ ] yes [ ] no _____
   Initials

2. I grant the investigator permission to share **with other researchers in the field** excerpts out of the original audio and video recordings and the transcribed data. Any personal or sensitive information will be removed, and any name or place references will be changed.

   [ ] yes [ ] no _____
   Initials

3. I grant the investigator permission to include excerpts out of the original audio and video recordings and the transcribed data **in an electronic questionnaire**. Any personal or sensitive information will be removed, and any name or place references will be changed.

   [ ] yes [ ] no _____
   Initials

Debriefing: Social Actions in German Conversation

Dear Participant:

Thank you very much for your participation in our study. Please note that you have the right to withdraw from this research project even after the study is completed. You can request that the recordings in which you participated may not be used.

When signing the consent form, you were informed that this study focuses on social actions in conversational German. This general research focus was given so that knowledge about the specific details of what we are investigating would not influence your interactions during the recording of the conversations. The precise nature of the study is the following:

This study investigates the ways in which humor is marked in every day conversation, often below the conscious awareness of the participants. Although humor is a phenomenon experienced and utilized by speakers on a daily basis, less is known about linguistic forms it takes. In order to utilize humor successfully, the speaker must mark it as such in order for other participants in the conversation to be able to react appropriately. The goal of this research is to reveal how speakers mark humor, with an emphasis the phonetic and phonological aspects of language and the reactions of other speakers in the conversation. This research hopes to contribute to the understanding of humor and how it appears and is utilized in everyday

conversation.

Please feel free to ask any further questions you may have.

Brandy Trygstad (trygstad@wisc.edu)
Joseph Salmons (jsalmons@wisc.edu)

Appendix 2.1.2. Conversational Data Collection Phase (German)

Einwilligungserklärung: **Soziales Verhalten in deutscher Konversation**

Sie werden hiermit eingeladen, an einer Studie teilzunehmen, die soziales Verhalten in gesprochenem Deutsch in Deutschland untersucht. Wir heißen Joseph Salmons und Brandy Trygstad. Joseph Salmons ist Professor und Brandy Trygstad ist Studentin in der Deutschabteilung der Universität von Wisconsin-Madison. Sie wurden als mögliche(r) Teilnehmer(in) ausgewählt, weil Sie Deutsch sprechen und mindestens 18 Jahre alt sind. Sie sind eine(r) von bis zu 50 Teilnehmern an dieser Studie.

Wenn Sie sich dafür entscheiden, an dieser Studie teilzunehmen, werden wir Sie und eine zweite Person Ihrer Wahl (z.B., eine(n) Freund(in) bzw. ein Familienmitglied) auf Audio und Video aufnehmen, während Sie alltägliche Konversationen durchführen, wie zum Beispiel Telefongespräche, Abendessen oder gesellige Zusammenkünfte. Darüber hinaus würden Sie ein Skript im Sprachlabor vorlesen. Alle Gesprächspartner(innen) müssen ihr Einverständnis zur Teilnahme an der Studie durch die Unterschrift dieser Einwilligungserklärung bestätigen. Jede Aufnahme wird zwischen ein und drei Stunden Ihrer Zeit in Anspruch nehmen und wir werden Sie, sollten Sie sich entschließen, an dieser Studie teilzunehmen, drei Mal aufnehmen. Die erste Aufnahme findet am Ort Ihrer Wahl statt, die zweite und dritte allerdings an der Universität Potsdam (Campus Neues Palais). Nachdem alle Aufnahmen fertig sind, werden die Daten linguistisch analysiert. Zusätzlich werden ausgewählte Stellen der Videodaten möglicherweise in einen elektronischen Fragebogen eingebaut.

Die mit dieser Studie verbundenen Risiken sind minimal und vergleichbar mit den Risiken, die das tägliche Leben mit sich bringt. Da die Studie allerdings alltägliche Unterhaltungen untersucht, ist ein unangenehmes Gefühl, dass durch die Aufnahme persönlicher Unterhaltungen z.B. mit Ihren Freund(innen) oder Familienmitgliedern nicht auszuschließen. Außerdem ist es eventuell möglich, dass TeilnehmerInnen, die den elektronischen Fragebogen ausfüllen, Sie am Gesicht erkennen könnten. Aus der Teilnahme an dieser Studie entstehen Ihnen keine Kosten. Falls jegliche Form von Missbrauch bzw. Vernachlässigung bei Ihnen zu Hause beobachtet bzw. befürchtet wird, sind die Forscher eventuell dazu verpflichtet, dies zu melden.

Aus der Teilnahme an der Studie sind keine direkten Vorteile zu erwarten. Eine finanzielle Entlohnung für die Teilnahme an dieser Studie ist nicht vorgesehen. Wir hoffen jedoch, dass wir Sie für Ihre Teilnahme damit entlohnen können, dass wir zu Ihrem Verständnis der gesprochenen Sprache und insbesondere Ihrem Wissen darüber, wie sich Muttersprachler(innen) in Alltagsgesprächen unterhalten, beitragen werden. Wenn Sie eine Kopie unserer Studie erhalten möchten, können Sie sich mit Joseph Salmons oder Brandy Trygstad unter den unten

angegebenen Telefonnummern bzw. E-Mail-Adressen in Verbindung setzen.

Bitte beachten Sie, dass alle Informationen, die in dieser Studie gesammelt werden und die mit Ihnen in Verbindung gebracht werden können, absolut vertraulich behandelt und nur mit Ihrer ausdrücklichen Erlaubnis weitergegeben werden. Alle Video- und Audiokassetten werden mit einem Code versehen, so dass die Etiketten auf den Kassetten keine Informationen enthalten, die direkt mit Ihnen in Verbindung gebracht werden könnten. In Transkripten werden Pseudonyme und Codes verwendet und die Stellen der Videodaten, die in den elektronischen Fragebogen eingebaut werden, werden so ausgesucht, dass sie keine privaten oder besonders peinlichen Informationen beinhalten. Außerdem werden die Videodaten gegen Herunterladen von Dritten verschlüsselt. Somit bleibt Ihre Privatsphäre möglichst geschutzt. Weiterhin werden wir die Kassetten unter Verschluss bewahren. Das Abspielen der Kassetten und die Verwendung der gesammelten und transkribierten Daten wird ausschließlich zu Forschungs- und Analysezwecken erfolgen. Die Kassetten werden an Dritte (andere Forscher eingeschlossen) nicht ohne Ihre vorherige ausdrückliche schriftliche Erlaubnis weitergegeben werden (Sie können diese Erlaubnis unten erteilen). Sie haben das Recht, Kassetten und Verschriftlichungen der Interaktionen, an denen Sie teilnehmen, einzusehen und ihre Veröffentlichung zu untersagen.

Nach Beendigung der Analyse werden die Kassetten für weitere Forschungszwecke an einem sicheren Ort aufbewahrt. Ihr Einverständnis vorausgesetzt ist es möglich, dass kurze Datenauszüge aus dieser Studie in der Zukunft für Präsentationen und Publikationen benutzt werden. Dabei werden für alle verwendeten Daten Pseudonyme und Codes zum Schutz Ihrer Identität verwendet.

Die Entscheidung ob Sie an dieser Studie teilnehmen möchten oder nicht wird keinerlei Einfluss auf Ihre Beziehung zu uns oder zu der Universität von Wisconsin-Madison haben. Sie werden in keinster Weise gezwungen, an dieser Studie teilzunehmen. Es steht Ihnen frei, (a) zu jeder Zeit die Teilnahme abzubrechen, (b) zu verlangen, dass die Aufnahmegeräte abgestellt werden, (c) zu verlangen, dass ein bereits aufgenommenes Band gelöscht und damit von der Studie ausgeschlossen wird.

Ihre Teilnahme an dieser Studie ist freiwillig. Mit Ihrer Unterschrift bestätigen Sie, dass Sie die obenstehende Informationen gelesen und verstanden haben und sich bereit erklären, an dieser Studie teilzunehmen. Auch wenn Sie dieses Formular unterschrieben haben, können Sie sich immer noch jederzeit aus der Studie zurückziehen.

Falls Sie Fragen haben, können Sie diese nun stellen. Falls Sie später weitere Informationen brauchen sollten, sind wir gerne bereit, Ihnen diese zu geben. Sie können uns unter folgender Adresse erreichen: 818 Van Hise Hall, 1220 Linden Drive, Madison, WI 53706, USA; Telefon: +1 608 262 2192, E-Mail: jsalmons@wisc.edu, trygstad@wisc.edu. Falls Sie Fragen bezüglich Ihrer Rechte als Forschungsteilnehmer(in) haben, können Sie sich an die folgende Institution wenden: Institutional Review Board Office, 310 Lathrop Hall, Madison WI 53706, USA; Telefon (R-Gespräch): +1 608 263 2320, Fax: +1 608 890 3900, lmlarson@ls.wisc.edu.

_____

Name der Teilnehmerin / des Teilnehmers

_____ _____
Unterschrift der Teilnehmerin / des Teilnehmers Datum

_____ _____
Unterschrift der Leiterin der Studie Datum

Bitte beantworten Sie die untenstehenden drei Fragen, indem Sie bei "ja/nein" ein Kreuzchen machen und denn mit Ihren Initialen unterschreiben. Sie haben das Recht, Kassetten und Verschriftlichungen der Interaktionen, an denen Sie teilnehmen, einzusehen und ihre Veröffentlichung zu untersagen.

1. Ich erlaube der Leiterin der Studie, Auszüge der originellen Video- bzw. der Tonaufnahmen und der verschriftlichen Daten **bei professionellen Vorträgen und in Publikationen** zu benutzen. Alle persönlichen bzw. privaten Informationen werden aussortiert und alle Namen- und Ortsangaben werden geändert.

   [ ] ja [ ] nein _____
   Initialen

2. Ich erlaube der Leiterin der Studie, Auszüge der originellen Video- bzw. der Tonaufnahmen und der verschriftlichen Daten **mit Dritten (anderen Forschern in diesem Gebiet)** zu teilen. Alle persönlichen bzw. privaten Informationen werden aussortiert und alle Namen- und Ortsangaben werden geändert.

   [ ] ja [ ] nein _____
   Initialen

3. Ich erlaube der Leiterin der Studie, Auszüge der originellen Video- bzw. der Tonaufnahmen und der verschriftlichen Daten **in einen elektronischen Fragebogen** einzubauen. Alle persönlichen bzw. privaten Informationen werden aussortiert und alle Namen- und Ortsangaben werden geändert.

   [ ] ja [ ] nein _____
   Initialen

Debriefing: Soziales Verhalten in deutscher Konversation

Liebe Teilnehmerin:
Lieber Teilnehmer:

wir danken Ihnen dafür, dass Sie an dieser Studie teilgenommen haben. Bitte beachten Sie, dass Sie das Recht haben, auch nach Abschluss dieses Projektes von Ihrer Teilnahme zurückzutreten. Sie können verlangen, dass die Aufnahmen, an denen Sie beteiligt sind, nicht verwendet werden.

Als Sie die Einwilligungserklärung unterschrieben, wurde Ihnen mitgeteilt, dass diese Studie soziales Verhalten in gesprochenem Deutsch untersucht. Der Untersuchungsgegenstand wurde absichtlich sehr allgemein gefasst, um zu vermeiden, dass ein detailliertes Wissen um den Zweck

der Studie Ihr Verhalten bei der Aufnahme der Konversationen beeinflusst. Der genaue Schwerpunkt dieser Studie ist der folgende:

Diese Studie untersucht, wie Humor in Alltagsgesprächen markiert wird, oftmals ohne dass dies den jeweiligen Gesprächsteilnehmern (explizit) bewusst ist. Obwohl Humor ein Phänomen ist, das von allen im Alltag erfährt und benutzt ist, ist wenig bekannt über die linguistischen Formen, die Humor annehmen kann. Um Humor erfolgreich zu verwenden, muss der Sprecher es als solche markieren, damit andere Teilnehmer am Gespräch sich darauf passend reagieren können. Die Zielsetzung dieser Studie ist es zu enthüllen, wie Gesprächsteilnehmer Sprache benutzen, um Humor zu markieren; Schwerpunkte der Studie sind phonetische bzw. phonologische Merkmale des Gesprächs und die Reaktionen von anderen Teilnehmer am Gespräch. Diese Studie trägt zum weiteren/tieferem Verstehen von Humor und wie dieses in Alltagsgesprächen auftritt und benutzt werden kann.

Falls Sie weitere Fragen hierzu haben, können Sie sie nun gerne stellen.

Brandy Trygstad (trygstad@wisc.edu)
Joseph Salmons (jsalmons@wisc.edu)

Appendix 2.1.3. Survey Phase (English translation)

Consent Form: Social Actions in German Conversation

You are invited to participate in a study that investigates the social actions in conversational German as spoken in Germany. We are Joseph Salmons and Brandy Trygstad. Joseph Salmons is a professor and Brandy Trygstad is a graduate student in the Department of German at the University of Wisconsin-Madison. You were selected as a possible participant in this study because you are a speaker of Standard German and are over the age of 18. You will be one of up to 600 subjects chosen to participate in this study.

If you decide to participate in this study, you will fill out an electronic questionnaire in a location of your choosing. Completion of the questionnaire would take up to between 30 and 45 minutes of your time. While filling out the questionnaire, you will be asked to view video and listen to audio excerpts. Therefore, you should select a location with appropriate computer hard- and software.

The risks involved in this study are minimal and are comparable to risks in everyday life. Since this study investigates everyday conversations your reactions to them, it is possible that you might feel uncomfortable with answering some questions. If you share personal information in your answers, it may be the case that you will not remain anonymous to the researchers. Any such information will be removed and will not be shared with other researchers or published. There is no cost to you for participating in this project.

There will be no direct benefit to you for participating in this research. You will receive no payment for your participation in this study. However, this study hopes to provide you with an indirect benefit by contributing to our knowledge of how language is used and how native

speakers of German interact.

Please note that any information that is obtained in connection with this study and that can be identified with you will remain confidential. More specifically, all questionnaire data will be coded so that no personally identifying information is attached to the files. We will keep all questionnaires in a secure place such as a locked file cabinet and password protected computer files. All data collected will be used for research and data analysis purposes only. The data will not be released to anyone, including other researchers, without your consent (you can give your consent to this below). You have the right to review your questionnaire responses, and to prohibit disclosure of any responses you provide at any time.

Following analysis, we will keep the questionnaires in a secure place for further related and unrelated research purposes conducted by us. In the future, very brief excerpts of the written responses you provide might be used for research publications if you give your consent to this below. All data used for these purposes will be coded to ensure the protection of your identity.

Your decision whether or not to participate will not affect your future relations with us or the University of Wisconsin-Madison. You are under no obligation to participate in this study. You are free to (a) discontinue participation in the study at any time, (b) request that completion of a questionnaire cease at any, (c) request that an already recorded questionnaire be destroyed and thus excluded from the study, and (d) decline to answer a question on the questionnaire.

Your participation is completely voluntary. By clicking on "Next", you indicate that you have read and understood the information provided above and have decided to participate. You are free to withdraw your consent and discontinue your participation at any time after clicking "Next" without consequences.

If you have any questions now, please ask me. If you need any additional information later, please do not hesitate to contact us. You can reach us at: 818 Van Hise Hall, 1220 Linden Drive, Madison, WI 53706, USA; telephone (608) 262-2192; E-mail: jsalmons@wisc.edu, trygstad@wisc.edu. Should you have any questions regarding your rights as a research subject, you can contact the Institutional Review Board Office, 310 Lathrop Hall, Madison WI 53706, USA; telephone (collect): (608) 263-2320, fax: (608) 890-3900; Email: lmlarson@ls.wisc.edu.

You are welcome to request a copy of this text from the researchers.

Please click on "Next" only if you can answer "yes" to all of the following questions. Clicking on "Next" counts as your consent. You have the right to review the answers you gave on the questionnaire, and you have the right to prohibit disclosure of any portions of tapes or transcripts.

1. I grant the investigator permission to collect my **responses to a questionnaire**. Any personal or sensitive information will be removed, and any name or place references will be changed.

2. I grant the investigator permission to use excerpts of my written responses to a questionnaire

at **professional meetings and in professional publications**. Any personal or sensitive information will be removed, and any name or place references will be changed.

3. I grant the investigator permission to **share with other researchers** in the field excerpts of my written responses to a questionnaire. Any personal or sensitive information will be removed, and any name or place references will be changed.

Debriefing: Social Actions in German Conversation

Dear Participant:

Thank you very much for your participation in our study. Please note that you have the right to withdraw from this research project even after the study is completed. You can request that the questionnaire data which you provided may not be used.

When signing the consent form, you were informed that this study focuses on social actions in conversational German. This general research focus was given so that knowledge about the specific details of what we are investigating would not influence your responses on the questionnaire. The precise nature of the study is the following:

This study investigates the ways in which humor is marked in every day conversation, often below the conscious awareness of the participants. Although humor is a phenomenon experienced and utilized by speakers on a daily basis, less is known about linguistic forms it takes. In order to utilize humor successfully, the speaker must mark it as such in order for other participants in the conversation to be able to react appropriately. The goal of this research is to reveal how speakers mark humor, with an emphasis the phonetic and phonological aspects of language and the reactions of other speakers in the conversation. This research hopes to contribute to the understanding of humor and how it appears and is utilized in everyday conversation.

Please feel free to ask any further questions you may have.

Brandy Trygstad (trygstad@wisc.edu)
Joseph Salmons (jsalmons@wisc.edu)

Appendix 2.1.4. Survey Phase (German)

Einwilligungserklärung: **Soziales Verhalten in deutscher Konversation**

Sie werden hiermit eingeladen, an einer Studie teilzunehmen, die soziales Verhalten in gesprochenem Deutsch in Deutschland untersucht. Wir heißen Joseph Salmons und Brandy Trygstad. Joseph Salmons ist Professor und Brandy Trygstad ist Studentin in der Deutschabteilung der Universität von Wisconsin-Madison. Sie wurden als mögliche(r) Teilnehmer(in) ausgewählt, weil Sie Deutsch sprechen und mindestens 18 Jahre alt sind. Sie sind eine(r) von bis zu 600 Teilnehmern an dieser Studie.

Wenn Sie sich dafür entscheiden, an dieser Studie teilzunehmen, werden Sie einen elektronischen Fragebogen an einem Ort Ihrer Wahl ausfüllen. Die Ausfüllung des Fragebogens wird zwischen 30 und 45 Minuten Ihrer Zeit in Anspruch nehmen. Während Sie den Fragebogen ausfüllen, werden wir Sie darum bitten, Video- und Tonaufnahmen anzuschauen. Sie sollten sich deswegen einen Ort aussuchen, an dem es die passenden Geräte und notwendige Software gibt.

Die mit dieser Studie verbundenen Risiken sind minimal und vergleichbar mit den Risiken, die das tägliche Leben mit sich bringt. Da die Studie allerdings alltägliche Unterhaltungen und Ihre Reaktionen dazu untersucht, kann es natürlich sein, dass Sie es manchmal als unangenehm empfinden werden, einige Fragen zu beantworten. Falls Sie persönliche bzw. private Informationen in Ihren Antworten mitteilen, kann es dazu kommen, dass Sie den Forschern gegenüber nicht anonym bleiben. Solche Informationen werden aussortiert und nicht mit Dritten geteilt oder veröffentlicht. Aus der Teilnahme an dieser Studie entstehen Ihnen keine Kosten.

Aus der Teilnahme an der Studie sind keine direkten Vorteile zu erwarten. Eine finanzielle Entlohnung für die Teilnahme an dieser Studie ist nicht vorgesehen. Wir hoffen jedoch, dass wir Sie für Ihre Teilnahme indirekt damit entlohnen können, dass wir zu Ihrem Verständnis der gesprochenen Sprache und insbesondere Ihrem Wissen darüber, wie sich Muttersprachler in Alltagsgesprächen unterhalten, beitragen werden.

Bitte beachten Sie, dass alle Informationen, die in dieser Studie gesammelt werden und die mit Ihnen in Verbindung gebracht werden können, äußerst vertraulich behandelt werden. Alle Fragebogen werden mit einem Code versehen, so dass sie keine Informationen enthalten, die direkt mit Ihnen in Verbindung gebracht werden könnten. Weiterhin werden wir die Fragebogen unter Verschluss bzw. in passwortgeschützten Dateien aufbewahren. Die Verwendung der gesammelten Daten wird ausschließlich zu Forschungs- und Analysezwecken erfolgen. Die Daten werden an Dritte (andere Forscher eingeschlossen) nicht ohne Ihre vorherige Erlaubnis weitergegeben werden (Sie können diese Erlaubnis unten erteilen). Sie haben das Recht, Ihre Antworten auf dem Fragebogen einzusehen und ihre Veröffentlichung zu untersagen.

Nach Beendigung der Analyse werden die Fragebogen für weitere Forschungszwecke an einem sicheren Ort aufbewahrt. Es ist möglich, dass kurze geschriebene Datenauszüge aus dieser Studie in der Zukunft für Präsentationen und Publikationen benutzt werden. Dabei werden für alle verwendeten Daten Pseudonyme und Codes zum Schutz Ihrer Identität verwendet.

Die Entscheidung ob Sie an dieser Studie teilnehmen möchten oder nicht wird keinerlei Einfluss auf Ihre Beziehung zu uns oder zu der Universität von Wisconsin-Madison haben. Sie werden in keinster Weise gezwungen, an dieser Studie teilzunehmen. Es steht Ihnen frei, (a) zu jeder Zeit die Teilnahme abzubrechen, (b) zu verlangen, dass die Erfüllung des Fragebogens abgebrochen wird, (c) zu verlangen, dass ein bereits ausgefüllter Fragebogen gelöscht und damit von der Studie ausgeschlossen wird, und (d) eine Frage auf dem Fragebogen nicht zu beantworten.

Ihre Teilnahme an dieser Studie ist freiwillig. Sie bestätigen, indem Sie auf „Weiter“ klicken, dass Sie die obenstehende Informationen gelesen und verstanden haben und sich bereit erklären, an dieser Studie teilzunehmen. Auch wenn Sie auf „Weiter“ geklickt haben, können Sie sich

immer noch jederzeit aus der Studie zurückziehen.

Falls Sie Fragen haben, können Sie diese nun stellen. Falls Sie später weitere Informationen brauchen sollten, sind wir gerne bereit, Ihnen diese zu geben. Sie können uns unter folgender Adresse erreichen: 818 Van Hise Hall, 1220 Linden Drive, Madison, WI 53706, USA; Telefon: +1 608 262 2192, E-Mail: jsalmons@wisc.edu, trygstad@wisc.edu. Falls Sie Fragen bezüglich Ihrer Rechte als Forschungsteilnehmer(in) haben, können Sie sich an die folgende Institution wenden: Institutional Review Board Office, 310 Lathrop Hall, Madison WI 53706, USA; Telefon (R-Gespräch): +1 608 263 2320, Fax: +1 608 890 3900, lmlarson@ls.wisc.edu.

Sie dürfen gerne die Forscher nach einer Teilnehmerkopie dieses Textes fragen.

Bitte klicken Sie nur auf „Weiter", wenn Sie alle der folgenden Fragen mit „ja" beantworten können. Das Klicken auf „Weiter" gilt als Ihre Einwilligung. Sie haben das Recht, Ihre Antworten auf einem Fragebogen, den Sie ausgefüllt haben, einzusehen und ihre Veröffentlichung zu untersagen.

1. Ich erlaube der Leiterin der Studie, meine **Antworten auf dem Fragebogen** zu sammeln. Alle persönlichen bzw. privaten Informationen werden aussortiert und alle Namen- und Ortsangaben werden geändert.

2. Ich erlaube der Leiterin der Studie, Auszüge meiner Antworten auf dem Fragebogen bei **professionellen Vorträgen und in Publikationen** zu benutzen. Alle persönlichen bzw. privaten Informationen werden aussortiert und alle Namen- und Ortsangaben werden geändert.

3. Ich erlaube der Leiterin der Studie, Auszüge meiner Antworten auf dem Fragebogen mit **Dritten** (anderen Forschern in diesem Gebiet) zu teilen. Alle persönlichen bzw. privaten Informationen werden aussortiert und alle Namen- und Ortsangaben werden geändert.

Debriefing: Soziales Verhalten in deutscher Konversation

Liebe Teilnehmerin:
Lieber Teilnehmer:

wir danken Ihnen dafür, dass Sie an dieser Studie teilgenommen haben. Bitte beachten Sie, dass Sie das Recht haben, auch nach Abschluss dieses Projektes von Ihrer Teilnahme zurückzutreten. Sie können verlangen, dass die Daten vom Fragebogen, den Sie ausgefüllt haben, nicht verwendet werden.

Als Sie die Einwilligungserklärung unterschrieben, wurde Ihnen mitgeteilt, dass diese Studie soziales Verhalten in gesprochenem Deutsch untersucht. Der Untersuchungsgegenstand wurde absichtlich sehr allgemein gefasst, um zu vermeiden, dass ein detailliertes Wissen um den Zweck der Studie Ihre Antworte auf den Fragebogen beeinflusst. Der genaue Schwerpunkt dieser Studie ist der folgende:

Diese Studie untersucht, wie Humor in Alltagsgesprächen markiert wird, oftmals ohne dass dies den jeweiligen Gesprächsteilnehmern (explizit) bewusst ist. Obwohl Humor ein Phänomen ist, das von allen im Alltag erfährt und benutzt ist, ist wenig bekannt über die linguistischen Formen, die Humor annehmen kann. Um Humor erfolgreich zu verwenden, muss der Sprecher es als solche markieren, damit andere Teilnehmer am Gespräch sich darauf passend reagieren können. Die Zielsetzung dieser Studie ist es zu enthüllen, wie Gesprächsteilnehmer Sprache benutzen, um Humor zu markieren; Schwerpunkte der Studie sind phonetische bzw. phonologische Merkmale des Gesprächs und die Reaktionen von anderen Teilnehmer am Gespräch. Diese Studie trägt zum weiteren/tieferem Verstehen von Humor und wie dieses in Alltagsgesprächen auftritt und benutzt werden kann.

Falls Sie weitere Fragen hierzu haben, können Sie sie nun gerne stellen.

Brandy Trygstad (trygstad@wisc.edu)
Joseph Salmons (jsalmons@wisc.edu)

Appendix 2.2. Data Collection Advertisements

Appendix 2.2.1. Conversational Data Collection Phase Recruitment Flyers and E-Mails

Note: The same text used in the below flyers was used in e-mails and Facebook messages sent to known contacts, with a request to forward to anybody who might have interest in participating.

Call for Participants in a Linguitics Study

Hello,

I am completing a dissertation in German linguistics at the University of Wisconsin-Madison, USA, and I am currently seeking native speakers to particpate in my study "Social Actions in German Conversation. Participation consists of holding a conversation of approximately 60 minutes duration with a one of your friends without restrictions on topic. This normal everyday conversation will be filmed with a video camera. All recordings will take place either at your home or at the University Potsdam and will each require at most 90 minutes of your time. All data will only be used for research purposes and will be used with utmost confidentially.

If you have any questions about the study and/or are interested in participating, please contact me by e-mail by 30 April. I would be happy to hear from you and am of course available to address your questions.

Sincerely,

Brandy Trygstad, M.A.

Contact: trygstad@wisc.edu

Linguistics Study
Brandy Trygstad
trygstad@wisc.edu
Linguistics Study
Brandy Trygstad
trygstad@wisc.edu
Linguistics Study
Brandy Trygstad
trygstad@wisc.edu
Linguistics Study
Brandy Trygstad
trygstad@wisc.edu
Linguistics Study
Brandy Trygstad
trygstad@wisc.edu
Linguistics Study
Brandy Trygstad
trygstad@wisc.edu
Linguistics Study
Brandy Trygstad
trygstad@wisc.edu
Linguistics Study
Brandy Trygstad
trygstad@wisc.edu
Linguistics Study
Brandy Trygstad
trygstad@wisc.edu
Linguistics Study
Brandy Trygstad
trygstad@wisc.edu
Linguistics Study
Brandy Trygstad
trygstad@wisc.edu
Linguistics Study
Brandy Trygstad
trygstad@wisc.edu

Aufruf: Teilnehmer an einer linguistischen Studie gesucht

Guten Tag,

ich promoviere in deutscher Sprachwissenschaft an der University of Wisconsin-Madison, USA, und ich suche muttersprachliche TeilnehmerInnen für meine Studie »Soziales Verhalten in deutscher Konversation«. Die Teilnahme besteht darin, mit einer Ihnen bekannten Person ein etwa 60-minütiges Gespräch ohne Themenvorgabe zu führen. Dieses normale Alltagsgespräch wird mit Videokamera gefilmt. Alle Aufnahmen finden entweder bei Ihnen zu Hause oder an der Uni Potsdam statt und beanspruchen jeweils maximal 90 Minuten Ihrer Zeit. Alle Daten werden nur zu Forschungszwecken verwendet und werden äußerst vertraulich behandelt.

Falls Sie Fragen zur Studie haben bzw. Interesse haben, daran teilzunehmen, melden Sie sich bitte bis zum 30.04. per E-Mail bei mir. Ich würde mich sehr freuen, Kontakt mit Ihnen aufzunehmen und bin gern bereit, Ihre Fragen zu beantworten.

Mit freundlichen Grüßen,

Brandy Trygstad, M.A.

Kontakt: trygstad@wisc.edu

Linguistische Studie
Brandy Trygstad
trygstad@wisc.edu

Appendix 2.2.2. Survey Phase Recruitment Flyers and E-Mails

Are you interested in German language?

If you are:
A native speaker of German
And over the age of 18,

you can participate in the study

Social Action
in German Conversation

(participate)!

For the study you only need:

1. A good internet connection in the location of your choice,
2. Headphones,
3. And approximately 30 minutes of time.

Enter the following web address and off you go!

http://go.wisc.edu/641152

* If you have any questions or concerns about the study, please feel free to contact the researcher, Brandy Trygstad (trygstad@wisc.edu), and/or her advisor, Joe Salmons (jsalmons@wisc.edu), at any time.

Interessieren Sie sich für die deutsche Sprache?

Wenn Sie:
Deutsch-Muttersprachler(in)
und über 18 Jahre alt sind,

können Sie an der Studie

Soziales Verhalten
in deutscher Konversation

teilnehmen!

Dafür brauchen Sie nur:

1. eine gute Internetverbindung am Ort Ihrer Wahl,
2. Kopfhörer
3. und ca. 30 Minuten Zeit.

Geben Sie die folgende Webadresse ein und los gehts!

http://go.wisc.edu/641152

* Falls Sie Fragen bzw. Bedenken bzgl. der Studie haben, können Sie sich jederzeit an die Leiterin der Studie, Brandy Trygstad (trygstad@wisc.edu), bzw. an ihren Betreuer, Joe Salmons (jsalmons@wisc.edu), wenden.

Recruitment E-Mail (English translation)

Please Forward: Seeking Participants for a Linguistic Study

I am currently looking for native speakers of German who would be willing to fill out an online survey. I was wondering if you could forward on the email below to anybody who might be interested?

Thank you in advance, and please let me know if you have any questions.

Brandy

Hello,

I am a PhD student in German Linguistics at the University of Wisconsin-Madison, USA. I am currently distributing a version of a survey, the results of which will serve as the basis of a portion of my dissertation research. I am looking for native speakers of German who would be willing to fill out the electronic questionnaire, which should take approximately 30 minutes. The link to the survey is here:

http://go.wisc.edu/641152

In order to complete the survey, it is best to be in a comfortable location with a good Internet connection and headphones.

If you are a native speaker of German over the age of 18, you are eligible to participate in the study. I invite you to do so now by clicking on the link above. Also, I invite you to forward this email to other native speakers of German who you might know. Please feel free to distribute this message widely.

This research has been approved by the IRB at the University of Wisconsin-Madison. Should you have any questions or concerns, please do not hesitate to contact me (trygstad@wisc.edu) or my dissertation advisor, Prof. Joe Salmons (jsalmons@wisc.edu), directly. We are happy to help at any time.

Sincerely,

Brandy Trygstad

Recruitment E-Mail (German)

Bitte weiterleiten: Teilnehmer(innen) für eine linguistische Studie gesucht

Ich suche derzeit Deutsch-Muttersprachler(innen), die dazu bereit wären, eine Online-Umfrage auszufüllen. Könntest Du bitte die folgende Mail an alle, die vielleicht daran Interesse hätten, weiterleiten?

Vielen Dank im Voraus. Sag mir bitte Bescheid, falls du Fragen dazu hast.

Brandy

Hallo,

ich promoviere in deutscher Sprachwissenschaft an der University of Wisconsin-Madison, USA. Ich verteile zur Zeit eine Umfrage, auf deren Ergebnissen ein Teil meiner Dissertation basiert ist. Ich suche deswegen Deutsch-Muttersprachler(innen), die bereit wären, an einer elektronischen Umfrage per Internet teilzunehmen. Dafür braucht man ungefähr 30 Minuten Zeit . Der Link zur

Umfrage steht unten:

http://go.wisc.edu/641152

Um den Fragebogen auszufüllen, sollte man sich am besten einen gemütlichen Ort mit einer guten Internetverbindung und Kopfhörern aussuchen.

Wenn Sie Deutsch-Muttersprachler(in)und über 18 Jahre alt sind, sind Sie qualifiziert und berechtigt, an dieser Studie teilzunehmen. Ich lade Sie herzlich ein, daran teilzunehmen, indem Sie den obigen Link anklicken. Ich lade Sie auch herzlich ein, diese Mail an andere Deutsch-Muttersprachler(innen), die Sie eventuell kennen, weiterzuleiten. Sie dürfen diese Nachricht gerne überall verteilen.

Diese Forschung ist von dem IRB an der University of Wisconsin-Madison anerkannt. Falls Sie dazu Fragen bzw. Bedenken haben, zögern Sie bitte nicht, sich an mich (trygstad@wisc.edu) oder an meinen Betreuer, Prof. Joe Salmons (jsalmons@wisc.edu) zu wenden. Wir stehen Ihnen gerne zur Verfügung.

Mit freundlichen Grüßen,

Brandy Trygstad

Appendix 2.3. Reading Script for Third Recording Session

Appendix 2.3.1. Reading Script (English translation)

Please read the following sentences aloud in a completely natural manner.

1. It was like Mr. Lehmann.
2. Also signed, I hope?
3. That is the gayest lighter I have ever seen.
4. You know that we are speaking on tape?
5. More important anyway is, can I a bum a smoke?
6. Who goes to campus in flip flops?
7. You're too assimilated, that's your problem.
8. Oh, crap.
9. The dog ate my homework.
10. When are we leaving finally?

Please read the following sentences aloud as if they are funny and/or with a smile.

1. It was like Mr. Lehmann.
2. Also signed, I hope?
3. That is the gayest lighter I have ever seen.
4. You know that we are speaking on tape?
5. More important anyway is, can I a bum a smoke?
6. Who goes to campus in flip flops?
7. You're too assimilated, that's your problem.
8. Oh, crap.
9. The dog ate my homework.
10. When are we leaving finally?

Appendix 2.3.2. Reading Script (German)

Lies bitte die folgenden Sätze ganz natürlich vor.

1. Das war so wie Herr Lehmann.
2. Auch unterschrieben, hoffe ich?

3. Das ist das schwulste Feuerzeug, was ich je gesehen habe.
4. Du weißt, dass wir sprechen auf einem Tonband?
5. Viel wichtiger ist sowieso, kann ich mir eine Kippe schnorren?
6. Wer geht denn in Flip Flops an die Uni?
7. Du bist zu angepasst, das ist dein Problem.
8. Ach, du Scheiße.
9. Der Hund hat meine Hausaufgaben gefressen.
10. Wann fahren wir endlich los?

Lies bitte jetzt die folgenden Sätze vor, als ob die lustig sind bzw. mit einem Lächeln.

1. Das war so wie Herr Lehmann.
2. Auch unterschrieben, hoffe ich?
3. Das ist das schwulste Feuerzeug, was ich je gesehen habe.
4. Du weißt, dass wir sprechen auf einem Tonband?
5. Viel wichtiger ist sowieso, kann ich mir eine Kippe schnorren?
6. Wer geht denn in Flip Flops an die Uni?
7. Du bist zu angepasst, das ist dein Problem.
8. Ach, du Scheiße.
9. Der Hund hat meine Hausaufgaben gefressen.
10. Wann fahren wir endlich los?

Appendix 2.4. Sample Survey Text

Appendix 2.4.1. Survey A1 (English translation)

Instructions

The following recordings contain every-day conversations among native speakers of German. Listen to and watch each recording **only once** and answer the corresponding questions. You may only look at the second questionnaire after you have answered all the questions.

Excerpt

Watch the recording **only once**!

((video embedded in the electronic questionnaire))

Questions

1. What will happen <u>next</u> in the conversation?

    - ☐ The participants will laugh.
    - ☐ The topic of conversation will change or was changed.
    - ☐ The conversation will simply continue, without any change.
    - ☐ Something else (please explain): *((open text field))*

2. Without regard to topic, how would you describe and/or characterize the contents of the conversation? Please mark <u>all responses</u> that match the excerpt.

    - ☐ An **answer** was given (to a question and/or an invitation)
    - ☐ An **invitation** was given
    - ☐ A **question** was asked
    - ☐ A **challenge** was made (a participant and/or a previous statement was challenged)
    - ☐ A **compliment** was given
    - ☐ A **criticism** was voiced (a participant and/or a previous statement was criticized)
    - ☐ The **topic** of conversation was changed
    - ☐ A **joke** was made
    - ☐ Something **else** (please explain): *((open text field))*

Appendix 2.4.2. Survey A1 (German)

Anleitungen

Die folgenden Aufnahmen beinhalten alltägliche Gespräche unter Muttersprachlern von Deutsch. Hören bzw. sehen Sie sich jede Aufnahme **nur ein Mal** an und beantworten Sie die entsprechenden Fragen. Erst nachdem Sie alle Fragen beantwortet haben, dürfen Sie den zweiten Fragebogen anschauen.

Ausschnitt

Schauen Sie sich die Aufnahmen **nur ein Mal** an!

((video embedded in the electronic questionnaire))

Fragen

1. Was wird als nächstes in dem Gespräch passieren?

    - ☐ Die Teilnehmer werden lachen.
    - ☐ Das Thema wird sich ändern bzw. wurde geändert.
    - ☐ Das Gespräch wird einfach weitergeführt werden, ohne jegliche Veränderung.
    - ☐ Etwas anderes (erklären Sie bitte): *((open text field))*

2. Wie würden Sie den Inhalt des Gesprächs beschreiben bzw. charakterisieren? Bitte markieren Sie alle Beschreibungen, die zum Ausschnitt passen.

    - ☐ Eine **Antwort** wurde gegeben (auf eine Frage bzw. eine Einladung)
    - ☐ Eine **Einladung** wurde gemacht
    - ☐ Eine **Frage** wurde gestellt
    - ☐ Eine **Herausforderung** wurde gemacht (Ein Teilnehmer bzw. eine vorherige Äusserung wurde herausgefordert)
    - ☐ Ein **Kompliment** wurde gegeben
    - ☐ Eine **Kritik** wurde geäußert (Ein Teilnehmer bzw. eine vorherige Äusserung wurde kritisiert)
    - ☐ Das **Thema** des Gesprächs wurde geändert
    - ☐ Ein **Witz** wurde gemacht
    - ☐ Etwas **anderes** (erklären Sie bitte): *((open text field))*

Appendix 2.4.3. Survey A2 (English translation)

Instructions

Listen to and watch the recordings one more time. At the same time, read the accompaning transcripts and answer the corresponding questions. You may listen to each excerpt multiple times. Please do **not** look up your answers on the first questionnaire.

Excerpt

You may watch the recording **more than once**.

((video embedded in the electronic questionnaire))

Transcript

((loose transcript in non-specialized format, words only))

Questions

1. What will happen next in the conversation?

    - ☐ The participants will laugh.
    - ☐ The topic of conversation will change or was changed.
    - ☐ The conversation will simply continue, without any change.
    - ☐ Something else (please explain): *((open text field))*

2. Why do you believe this? *((open text field))*

3. Without regard to topic, how would you describe and/or characterize the contents of the lines which are indicated by the arrow? Please mark all responses that match the excerpt.

    - ☐ An **answer** was given (to a question and/or an invitation)
    - ☐ An **invitation** was given
    - ☐ A **question** was asked
    - ☐ A **challenge** was made (a participant and/or a previous statement was challenged)
    - ☐ A **compliment** was given
    - ☐ A **criticism** was voiced (a participant and/or a previous statement was criticized)
    - ☐ The **topic** of conversation was changed
    - ☐ A **joke** was made
    - ☐ Something **else** (please explain): *((open text field))*

4. Why would you describe the contents of the lines indicated above in this way? *((open text field))*

5. Name and describe the important and/or noteworthy features of the conversation which you noticed in order to answer the above questions.

    1. *((open text field))*
    2. *((open text field))*
    3. *((open text field))*
    4. *((open text field))*

Appendix 2.4.4. Survey A2 (German)

Anleitungen

Hören bzw. sehen Sie sich nun die Aufnahmen noch einmal an. Lesen Sie dabei die dazugehörenden Transkripte und beantworten Sie die entsprechenden Fragen. Sie dürfen sich jeden Ausschnitt mehrmals anhören bzw. ansehen. Bitte schlagen Sie Ihre Antworten auf dem

ersten Fragebogen **nicht** nach.

Ausschnitt

Sie dürfen sich die Aufnahme **mehrmals** anschauen.

((video embedded in the electronic questionnaire))

Transkript

((loose transcript in non-specialized format, words only))

Fragen

1. Was wird als nächstes in dem Gespräch passieren?

    - ☐ Die Teilnehmer werden lachen.
    - ☐ Das Thema wird sich ändern bzw. wurde geändert.
    - ☐ Das Gespräch wird einfach weitergeführt worden, ohne jegliche Veränderung.
    - ☐ Etwas anderes (erklären Sie bitte): *((open text field))*

2. Warum glauben Sie das? *((open text field))*

3. Abgesehen vom Thema, wie würden Sie den Inhalt der Zeile, auf die der Pfeil zeigt, beschreiben bzw. charakterisieren? Bitte markieren Sie alle Beschreibungen, die zum Ausschnitt passen.

    - ☐ Eine **Antwort** wurde gegeben (auf eine Frage bzw. eine Einladung)
    - ☐ Eine **Einladung** wurde gemacht
    - ☐ Eine **Frage** wurde gestellt
    - ☐ Eine **Herausforderung** wurde gemacht (Ein Teilnehmer bzw. eine vorherige Äusserung wurde herausgefordert)
    - ☐ Ein **Kompliment** wurde gegeben
    - ☐ Eine **Kritik** wurde geäußert (Ein Teilnehmer bzw. eine vorherige Äusserung wurde kritisiert)
    - ☐ Das **Thema** des Gesprächs wurde geändert
    - ☐ Ein **Witz** wurde gemacht
    - ☐ Etwas **anderes** (erklären Sie bitte): *((open text field))*

4. Warum würden Sie den Inhalt der obengenannten Zeile so beschreiben? *((open text field))*

5. Nennen und beschreiben Sie die wichtigen bzw. auffälligen Merkmale des Gesprächs, auf die Sie geachtet haben, um die obigen Fragen zu beantworten:

    1. *((open text field))*

2. *((open text field))*
3. *((open text field))*
4. *((open text field))*

Appendix 2.4.5. Survey B1 (English translation)

Instructions

The following transcripts contain every-day conversations among native speakers of German. Read each transcript and answer the corresponding questions. You may only look at the second questionnaire **after** you have answered all the questions.

Transcript

((loose transcript in non-specialized format, words only))

Questions

1. What will happen next in the conversation?

   - ☐ The participants will laugh.
   - ☐ The topic of conversation will change or was changed.
   - ☐ The conversation will simply continue, without any change.
   - ☐ Something else (please explain): *((open text field))*

2. Why do you believe this? *((open text field))*

3. Without regard to topic, how would you describe and/or characterize the contents of the lines which are indicated by the arrow? Please mark all responses that match the excerpt.

   - ☐ An **answer** was given (to a question and/or an invitation)
   - ☐ An **invitation** was given
   - ☐ A **question** was asked
   - ☐ A **challenge** was made (a participant and/or a previous statement was challenged)
   - ☐ A **compliment** was given
   - ☐ A **criticism** was voiced (a participant and/or a previous statement was criticized)
   - ☐ The **topic** of conversation was changed
   - ☐ A **joke** was made
   - ☐ Something **else** (please explain): *((open text field))*

4. Why would you describe the contents of the lines indicated above in this way? *((open text field))*

5. Name and describe the important and/or noteworthy features of the conversation which you noticed in order to answer the above questions.

1. *((open text field))*
2. *((open text field))*
3. *((open text field))*
4. *((open text field))*

Appendix 2.4.6. Survey B1 (German)

Anleitungen

Die folgenden Transkripte beinhalten alltägliche Gespräche unter Muttersprachlern von Deutsch. Lesen Sie jedes Transkript und beantworten Sie die entsprechenden Fragen. **Erst nachdem** Sie alle Fragen beantwortet haben, dürfen Sie den zweiten Fragebogen anschauen.

Transkript

((loose transcript in non-specialized format, words only))

Fragen

1. Was wird als nächstes in dem Gespräch passieren?

   - ☐ Die Teilnehmer werden lachen.
   - ☐ Das Thema wird sich ändern bzw. wurde geändert.
   - ☐ Das Gespräch wird einfach weitergeführt worden, ohne jegliche Veränderung.
   - ☐ Etwas anderes (erklären Sie bitte): *((open text field))*

2. Warum glauben Sie das? *((open text field))*

3. Abgesehen vom Thema, wie würden Sie den Inhalt der Zeile, auf die der Pfeil zeigt, beschreiben bzw. charakterisieren? Bitte markieren Sie alle Beschreibungen, die zum Ausschnitt passen.

   - ☐ Eine **Antwort** wurde gegeben (auf eine Frage bzw. eine Einladung)
   - ☐ Eine **Einladung** wurde gemacht
   - ☐ Eine **Frage** wurde gestellt
   - ☐ Eine **Herausforderung** wurde gemacht (Ein Teilnehmer bzw. eine vorherige Äusserung wurde herausgefordert)
   - ☐ Ein **Kompliment** wurde gegeben
   - ☐ Eine **Kritik** wurde geäußert (Ein Teilnehmer bzw. eine vorherige Äusserung wurde kritisiert)
   - ☐ Das **Thema** des Gesprächs wurde geändert
   - ☐ Ein **Witz** wurde gemacht
   - ☐ Etwas **anderes** (erklären Sie bitte): *((open text field))*

4. Warum würden Sie den Inhalt der obengenannten Zeile so beschreiben? *((open text field))*

5. Nennen und beschreiben Sie die wichtigen bzw. auffälligen Merkmale des Gesprächs, auf die Sie geachtet haben, um die obigen Fragen zu beantworten:

    1. *((open text field))*
    2. *((open text field))*
    3. *((open text field))*
    4. *((open text field))*

Appendix 2.4.7. Survey B2 (English translation)

Instructions

The following recordings contain every-day conversations among native speakers of German. Listen to and watch the recordings. At the same time, read the accompanying transcripts and answer the corresponding questions. You may listen to each excerpt multiple times. Please do **not** look up your answers on the first questionnaire.

Excerpt

You may watch the recording **more than once**.

((video embedded in the electronic questionnaire))

Transcript

((loose transcript in non-specialized format, words only))

Questions

1. What will happen next in the conversation?

    - ☐ The participants will laugh.
    - ☐ The topic of conversation will change or was changed.
    - ☐ The conversation will simply continue, without any change.
    - ☐ Something else (please explain): *((open text field))*

2. Why do you believe this? *((open text field))*

3. Without regard to topic, how would you describe and/or characterize the contents of the lines which are indicated by the arrow? Please mark all responses that match the excerpt.

    - ☐ An **answer** was given (to a question and/or an invitation)
    - ☐ An **invitation** was given
    - ☐ A **question** was asked
    - ☐ A **challenge** was made (a participant and/or a previous statement was challenged)
    - ☐ A **compliment** was given

- ☐ A **criticism** was voiced (a participant and/or a previous statement was criticized)
- ☐ The **topic** of conversation was changed
- ☐ A **joke** was made
- ☐ Something **else** (please explain): *((open text field))*

4. Why would you describe the contents of the lines indicated above in this way? *((open text field))*

5. Name and describe the important and/or noteworthy features of the conversation which you noticed in order to answer the above questions.

    1. *((open text field))*
    2. *((open text field))*
    3. *((open text field))*
    4. *((open text field))*

Appendix 2.4.8. Survey B2 (German)

Anleitungen

Die folgenden Aufnahmen beinhalten alltägliche Gespräche unter Muttersprachlern von Deutsch. Hören bzw. sehen Sie sich die Aufnahmen an. Lesen Sie dabei die dazugehörenden Transkripte und beantworten Sie die entsprechenden Fragen. Sie dürfen sich jeden Ausschnitt mehrmals anhören. Bitte schlagen Sie Ihre Antworten auf dem ersten Fragebogen **nicht** nach.

Ausschnitt

Sie dürfen sich die Aufnahme **mehrmals** anschauen.

((video embedded in the electronic questionnaire))

Transkript

((loose transcript in non-specialized format, words only))

Fragen

1. Was wird als nächstes in dem Gespräch passieren?

    - ☐ Die Teilnehmer werden lachen.
    - ☐ Das Thema wird sich ändern bzw. wurde geändert.
    - ☐ Das Gespräch wird einfach weitergeführt worden, ohne jegliche Veränderung.
    - ☐ Etwas anderes (erklären Sie bitte): *((open text field))*

2. Warum glauben Sie das? *((open text field))*

3. Abgesehen vom Thema, wie würden Sie den Inhalt der Zeile, auf die der Pfeil zeigt, beschreiben bzw. charakterisieren? Bitte markieren Sie alle Beschreibungen, die zum Ausschnitt passen.

    - ☐ Eine **Antwort** wurde gegeben (auf eine Frage bzw. eine Einladung)
    - ☐ Eine **Einladung** wurde gemacht
    - ☐ Eine **Frage** wurde gestellt
    - ☐ Eine **Herausforderung** wurde gemacht (Ein Teilnehmer bzw. eine vorherige Äusserung wurde herausgefordert)
    - ☐ Ein **Kompliment** wurde gegeben
    - ☐ Eine **Kritik** wurde geäußert (Ein Teilnehmer bzw. eine vorherige Äusserung wurde kritisiert)
    - ☐ Das **Thema** des Gesprächs wurde geändert
    - ☐ Ein **Witz** wurde gemacht
    - ☐ Etwas **anderes** (erklären Sie bitte): *((open text field))*

4. Warum würden Sie den Inhalt der obengenannten Zeile so beschreiben? *((open text field))*

5. Nennen und beschreiben Sie die wichtigen bzw. auffälligen Merkmale des Gesprächs, auf die Sie geachtet haben, um die obigen Fragen zu beantworten:

    1. *((open text field))*
    2. *((open text field))*
    3. *((open text field))*
    4. *((open text field))*

Appendix 2.5. Demographics Survey Questions

Appendix 2.5.1. Demographics Survey Questions (English translation)

About You

Before you begin the questionnaire, please answer the following questions about yourself.

How old are you? *((open text field))*

Do you identify as:

- ☐ male
- ☐ female
- ☐ neither

Where are you originally from?

City *((open text field))*
State *((open text field))*
Country *((open text field))*

Where did you grow up?

City *((open text field))*
State *((open text field))*
Country *((open text field))*

Do you speak any dialects?

- ☐ yes
- ☐ no

If yes, which? *((open text field))*

What are you studying or what did you study at university?

Major 1 *((open text field))*
Major 2 *((open text field))*
Minor 1 *((open text field))*
Minor 2 *((open text field))*
Area of Study / Department *((open text field))*

In which semester of study are you? If you have already completed your studies, in what year did you graduate? *((open text field))*

What do you do for your career? If you are still studying at university, what do you plan to do for your career after graduation? *((open text field))*

Appendix 2.5.2. Demographics Survey Questions (German)

Über Sie

Bevor Sie mit dem Fragebogen anfangen, beantworten Sie bitte die folgenden Fragen über Sich selbst.

Wie alt sind Sie? *((open text field))*

Identifizieren Sie Sich als:

- ☐ Mann
- ☐ Frau
- ☐ weder noch

Woher kommen Sie ursprünglich?

Stadt *((open text field))*
Bundesland *((open text field))*
Land *((open text field))*

Wo sind Sie aufgewachsen?

Stadt *((open text field))*
Bundesland *((open text field))*
Land *((open text field))*

Sprechen Sie irgendwelche Dialekte?

- ☐ ja
- ☐ nein

Wenn ja, welchen/welche? *((open text field))*

Was studieren Sie bzw. was haben Sie studiert?

Hauptfach 1 *((open text field))*
Hauptfach 2 *((open text field))*
Nebenfach 1 *((open text field))*
Nebenfach 2 *((open text field))*
Fachgebiet bzw. Falkutät *((open text field))*

In welchem Semester sind Sie? Falls Sie bereits ein Studium abgeschlossen haben, in welchem Jahr haben Sie Ihren ersten Abschluss gemacht? *((open text field))*

Was machen Sie beruflich? Wenn Sie immer noch studieren, was haben Sie vor, nach Ihrem Abschluss beruflich zu machen? *((open text field))*

Appendix 3.1. Transcript Symbols

[ ]	overlapping talk
↑	rise in pitch
↓	drop in pitch
.	falling intonation
;	slightly falling intonation
,	continuing or slightly rising intonation
¿	rising intonation
?	sharply rising intonation
> <	quicker than surrounding talk; can be stacked >> << if a lot quicker
< >	slower than surrounding talk; can be stacked << >> if a lot slower
XX	capitalized syllables louder than surrounding talk
° °	quieter than surrounding talk
ʔ	glottal stop
·h	inhalation
h	exhalation
ptch	lip smack, audible parting of lips
:	syllable lengthening
–	syllable cut off before completion
=	latching of utterances without gap
(X.X)	timed pause; time indicated in seconds
( )	utterance unclear (if transcribed); utterance inaudible (if left empty)
☺	utterance treated as non-serious; “smile voice”
→	utterance of interest

Appendix 3.2. PRAAT Script by Nicholas Williams: beatnik_v1.0.2

```
# Beatnik

clearinfo

# threshold in db
# timestep in seconds
threshold=65
step = 0.01

objName$ = selected$ ("Sound")

To TextGrid... one bell

select Sound 'objName$'
Scale intensity... 70
dur = Get total duration

View & Edit
editor Sound 'objName$'

Show all
Move cursor to... 0.0
rightTime = 0
beats = 0

while rightTime < dur
     left = Get intensity
     leftTime = Get cursor

     Move cursor by... step
     center = Get intensity
     centerTime = Get cursor

     Move cursor by... step
     right = Get intensity
     rightTime = Get cursor

     if center = undefined or left = undefined or right =
undefined and center > threshold
     else

          if center > left
               if center > right
               beats = beats + 1
               time[beats] = centerTime
```

```
                printline peak: 'center' @ 'centerTime'
                endif
           endif

     endif
endwhile

Close

printline Beats: 'beats'

select TextGrid 'objName$'
for i to beats
     Insert boundary... 1 time[i]
endfor

plus Sound 'objName$'
View & Edit
```

Appendix 4.1. Perception Data: All Participants

Figure 4.17.1.

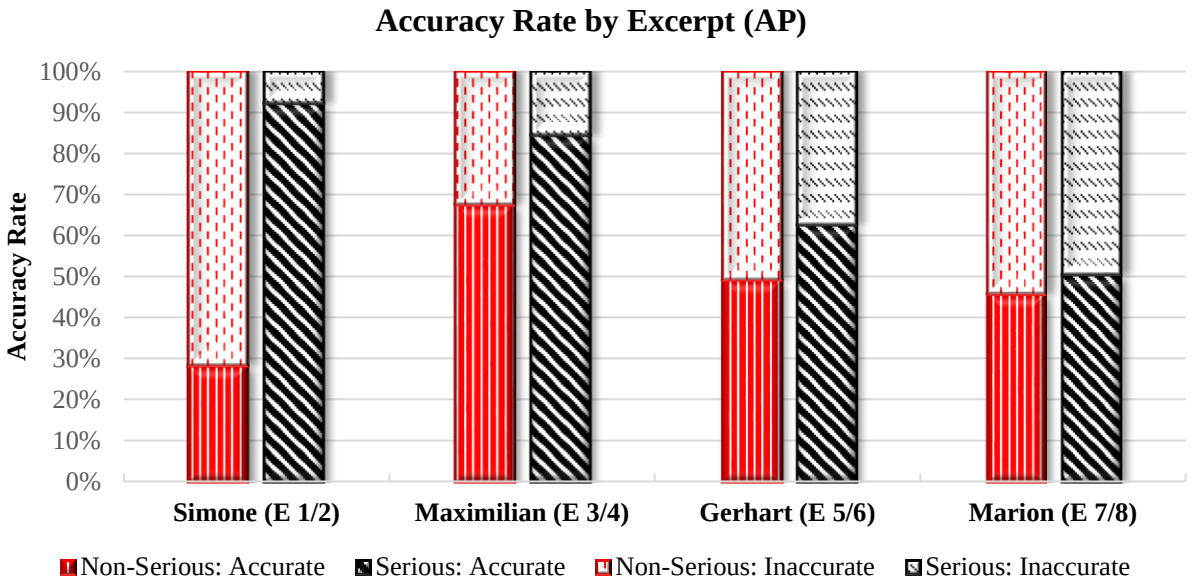


Figure 4.18.1.

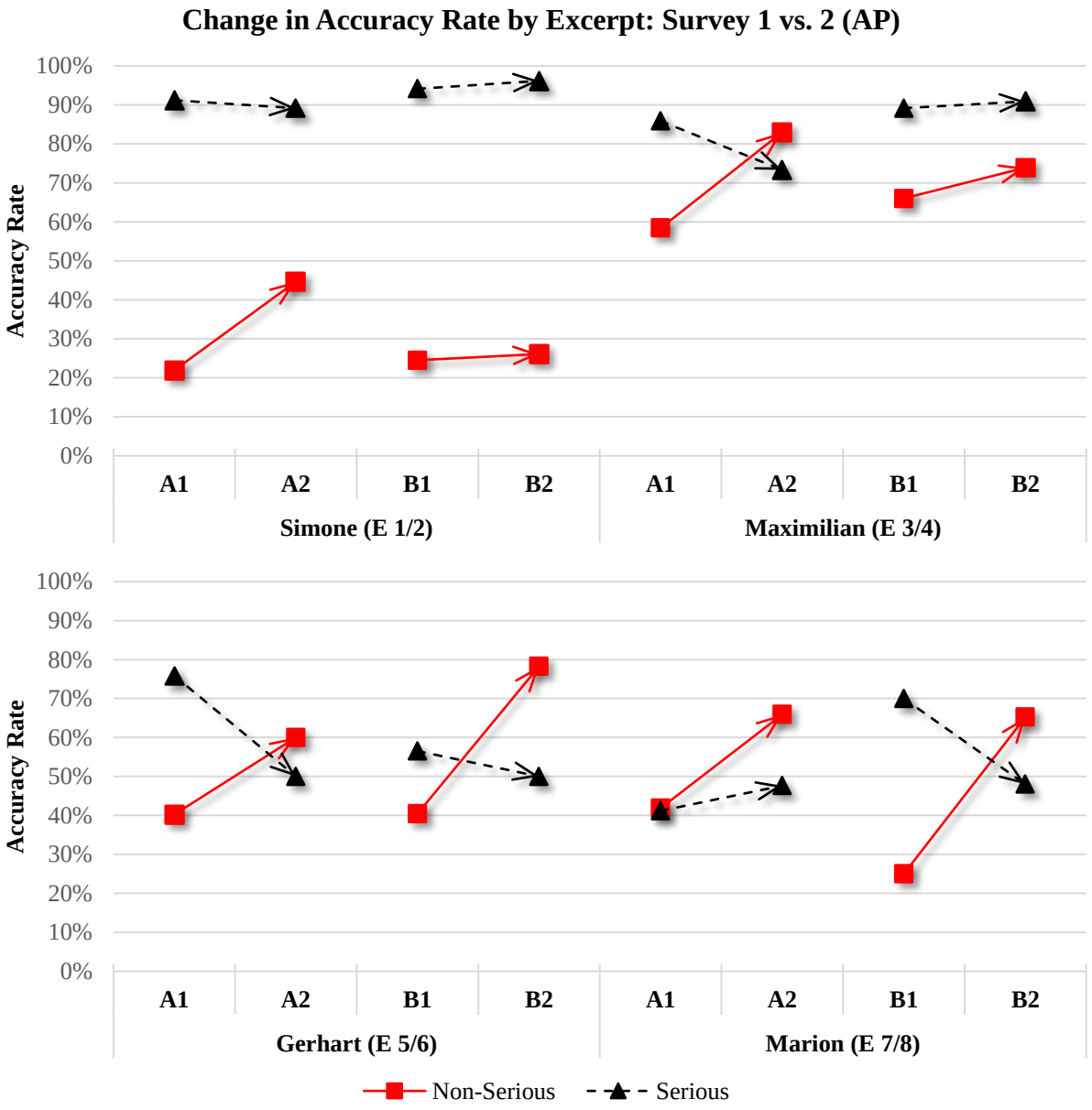

Change in Accuracy Rate by Excerpt: Survey 1 vs. 2 (AP)
Accuracy Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A1
A2
B1
B2
Simone (E 1/2)
Maximilian (E 3/4)
Gerhart (E 5/6)
Marion (E 7/8)
Non-Serious
Serious

Figure 4.28.1.

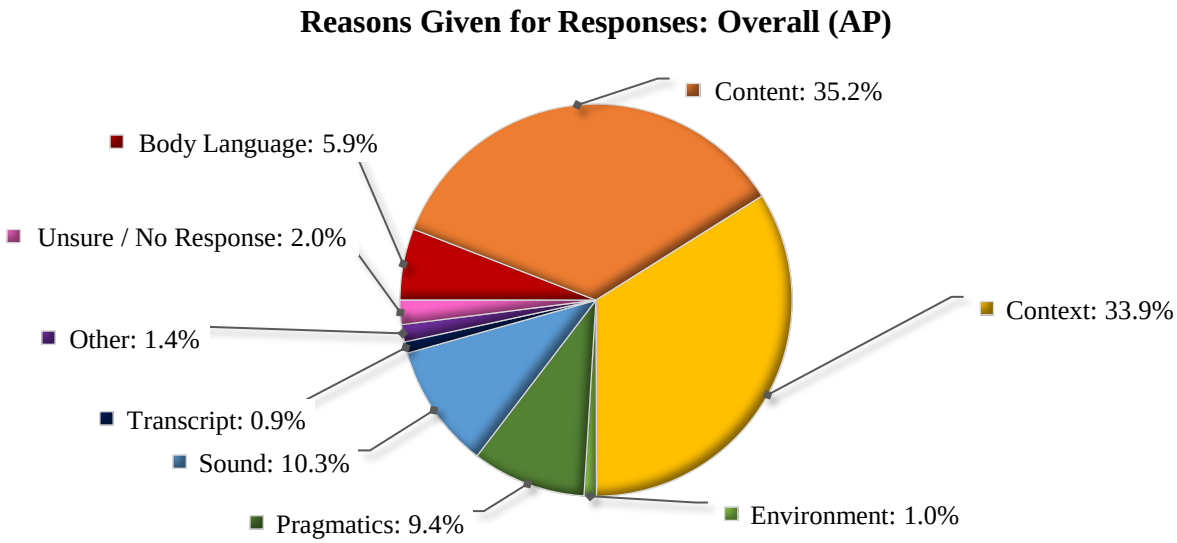


Figure 4.29.1.

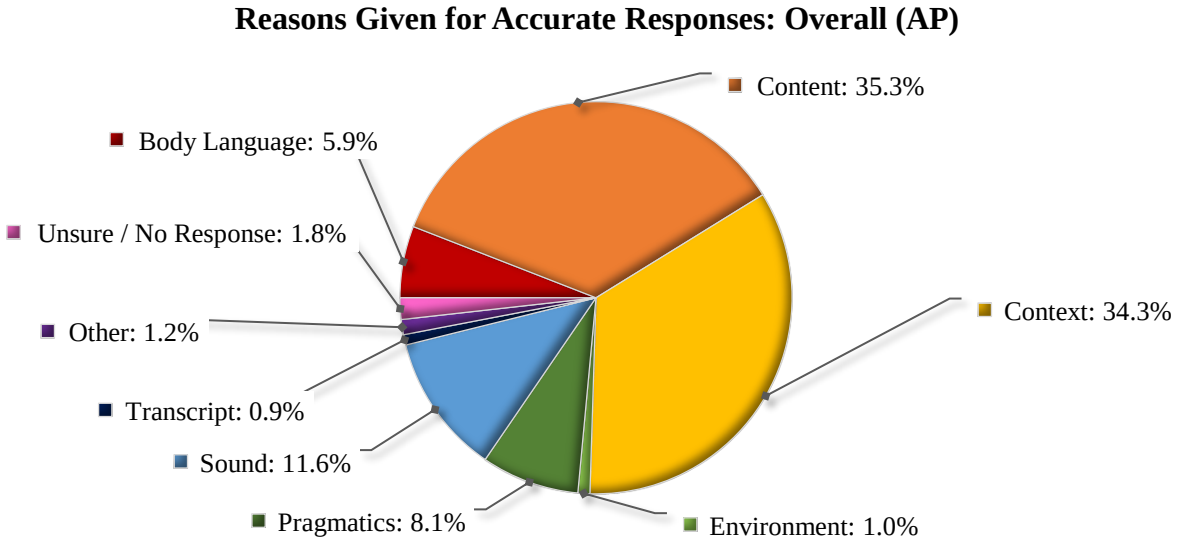


Figure 4.30.1.

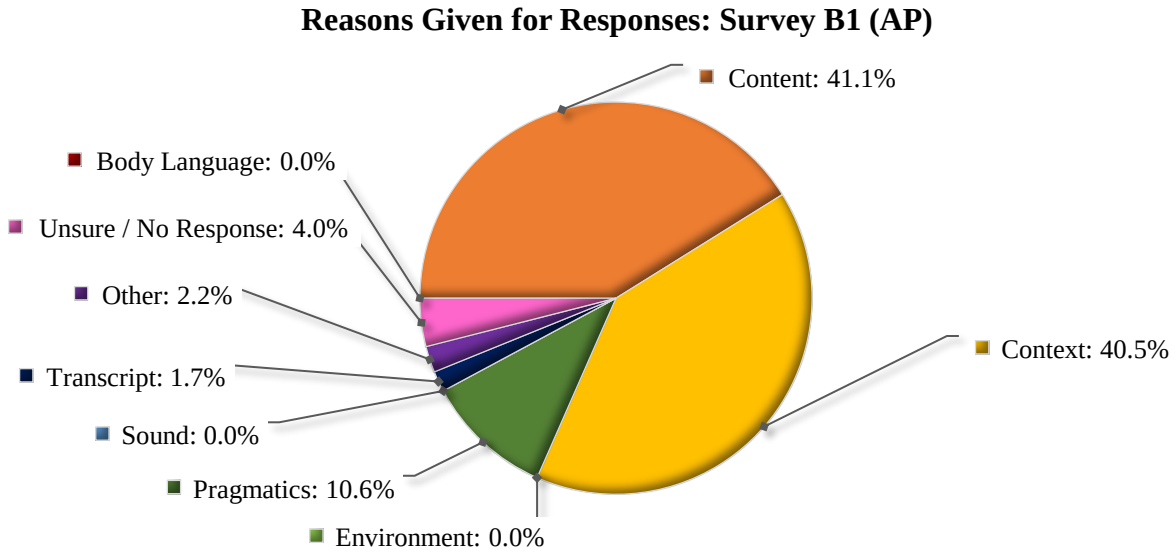


Figure 4.31.1.

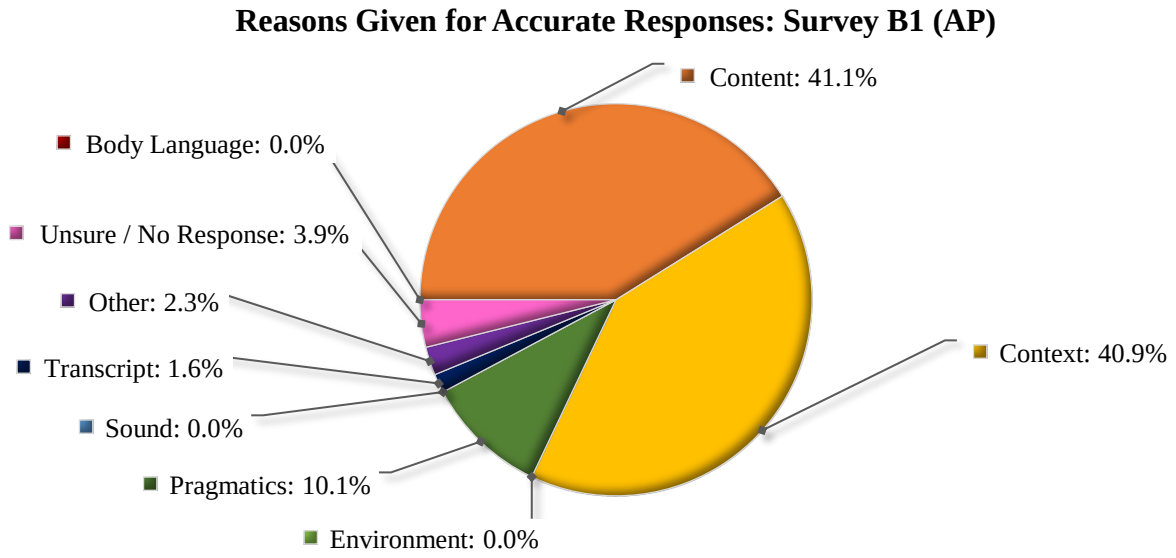


Figure 4.32.1.

Reasons Given for Responses: Surveys A2 and B2 (AP)

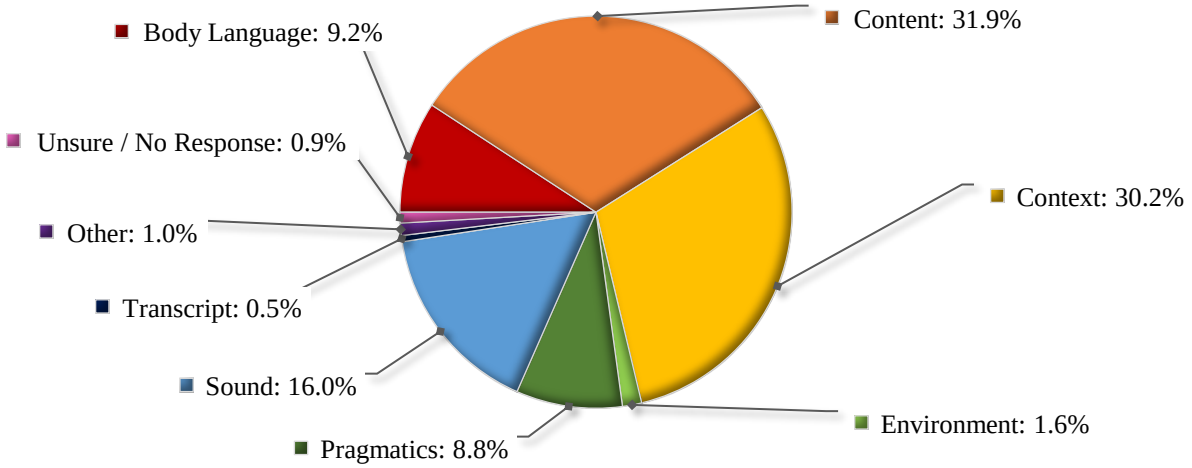


Figure 4.33.1.

Reasons Given for Accurate Responses: Surveys A2 and B2 (AP)

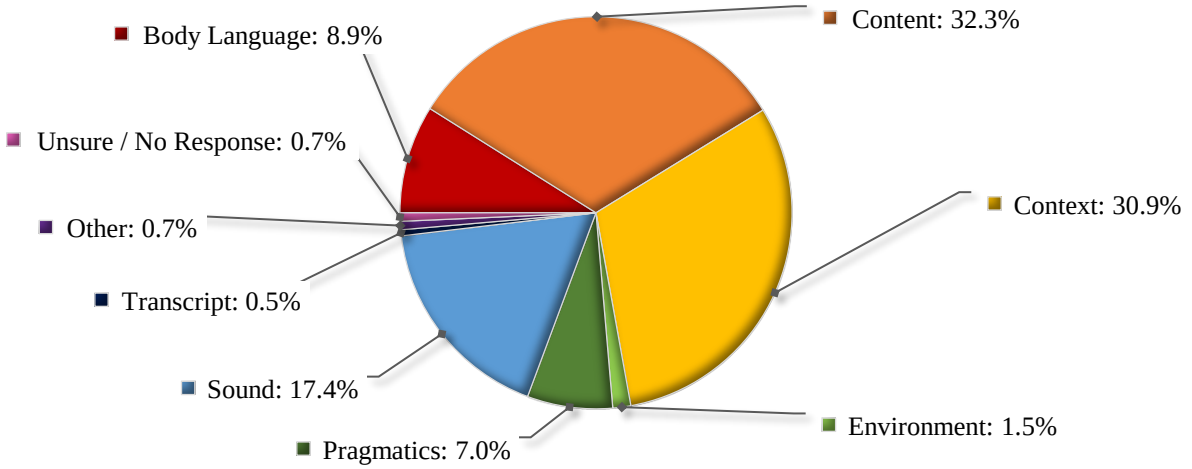


Figure 4.34.1.

Reasons Given for Responses: Survey A2 (AP)

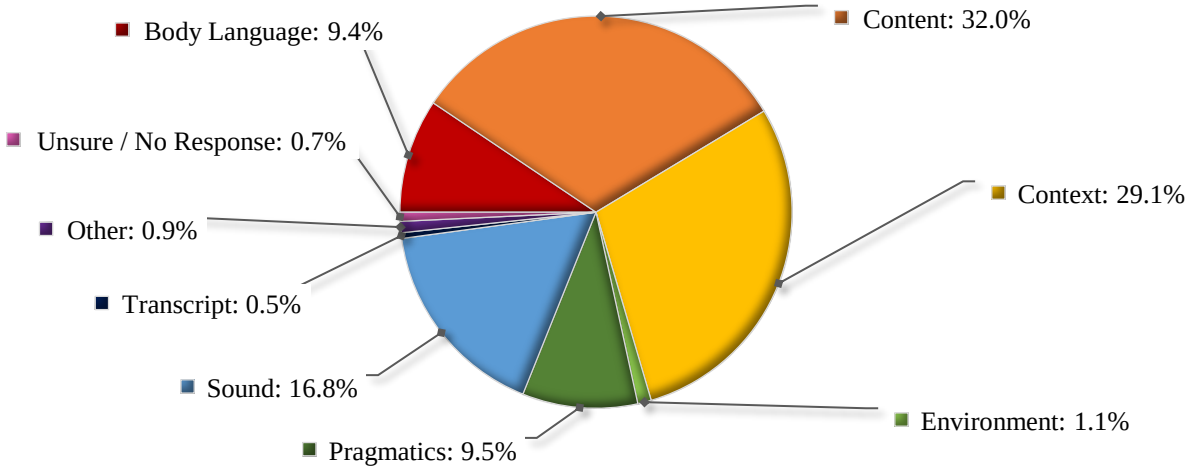


Figure 4.35.1.

Reasons Given for Accurate Responses: Survey A2 (AP)

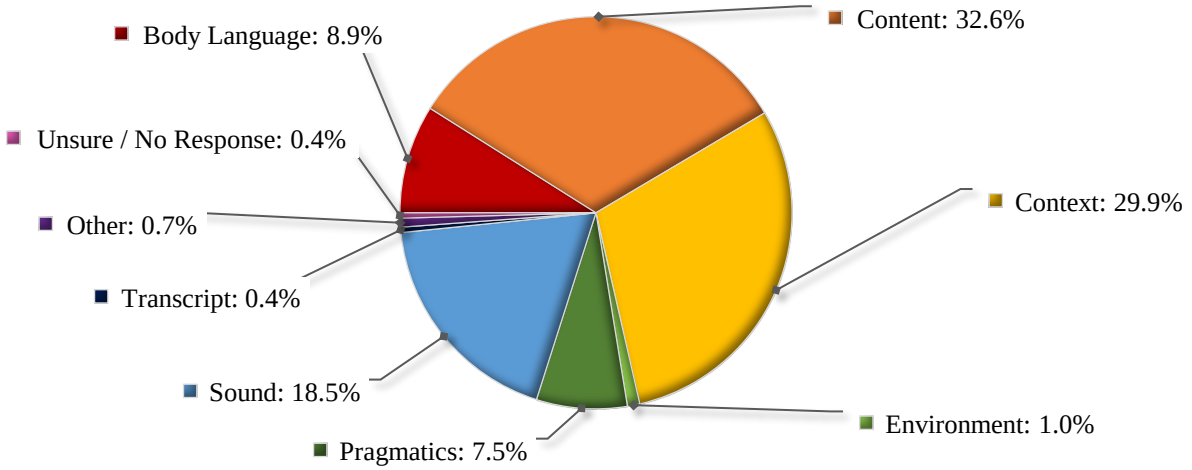


Figure 4.36.1.

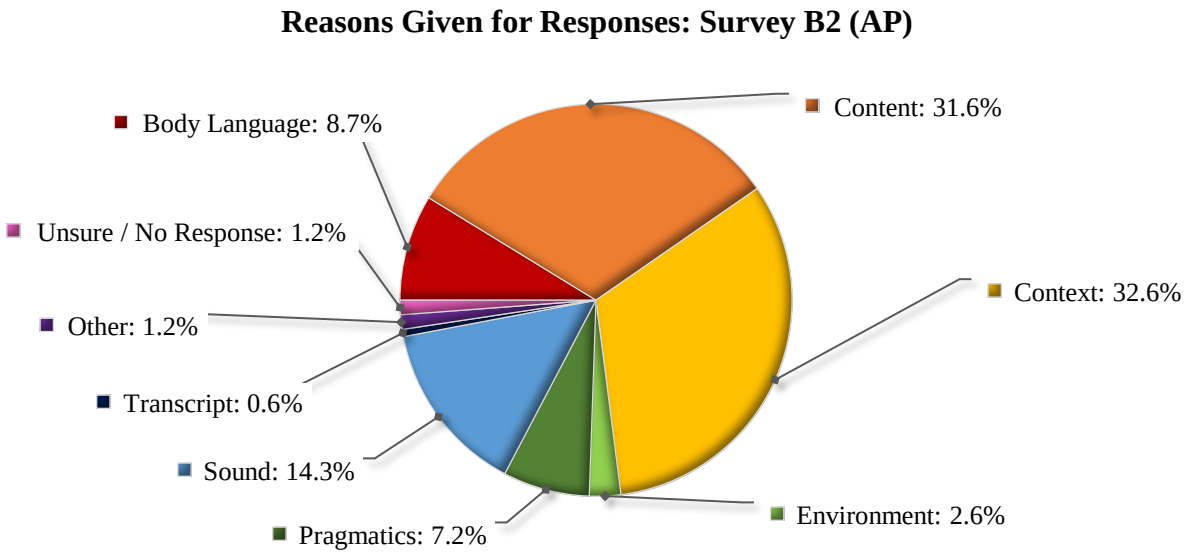


Figure 4.37.1.

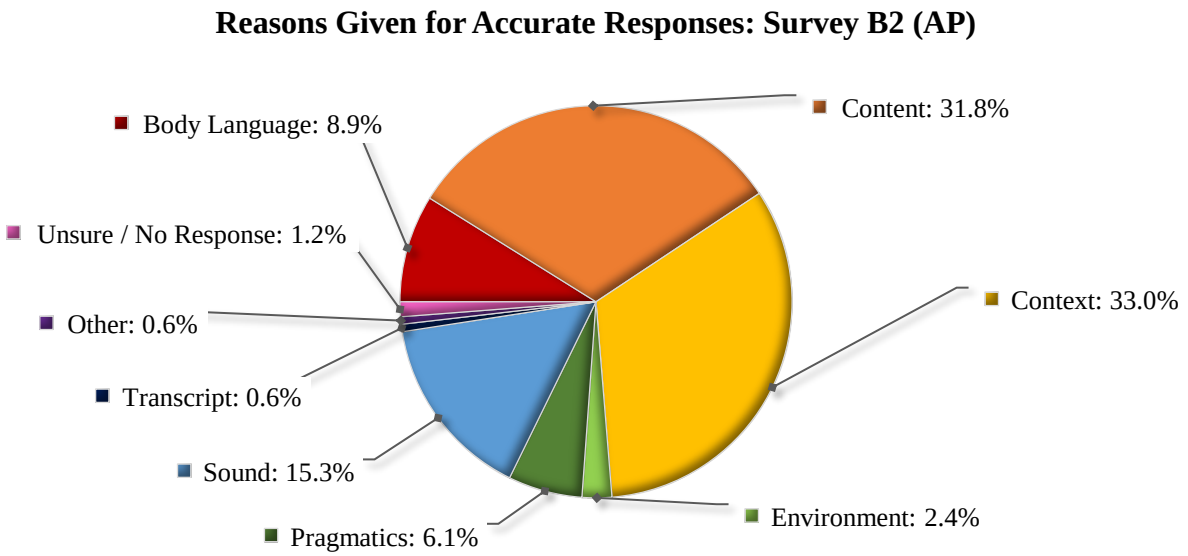


Figure 4.38.1.

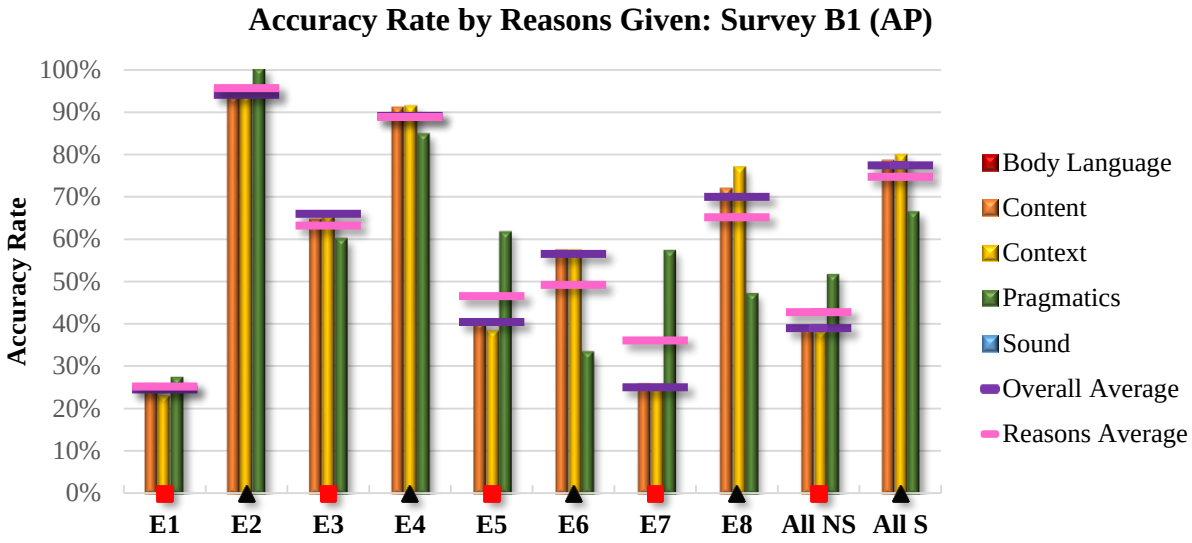


Figure 4.39.1.

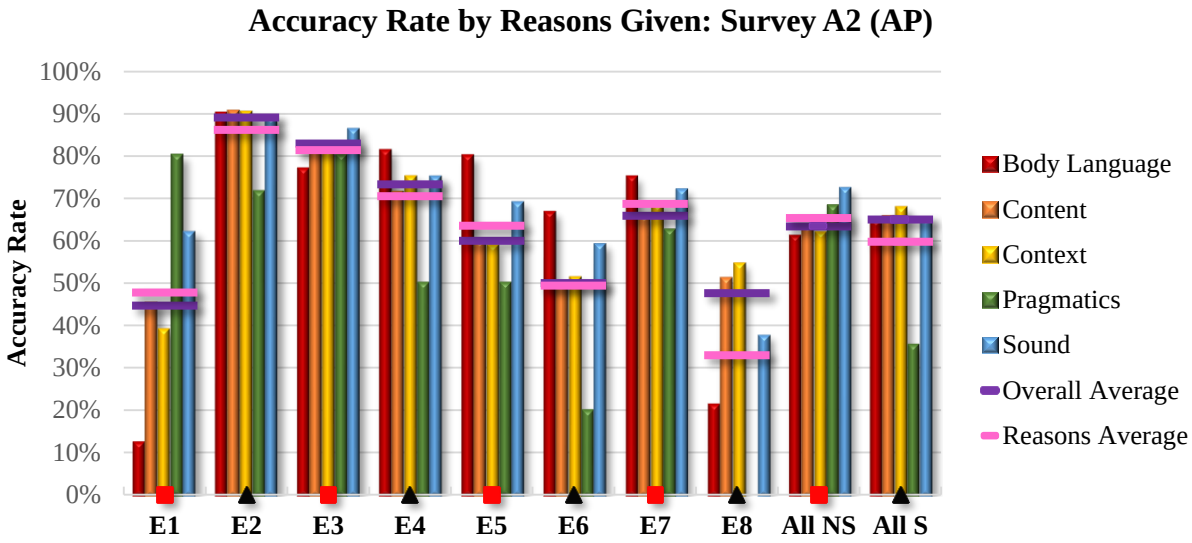


Figure 4.40.1.

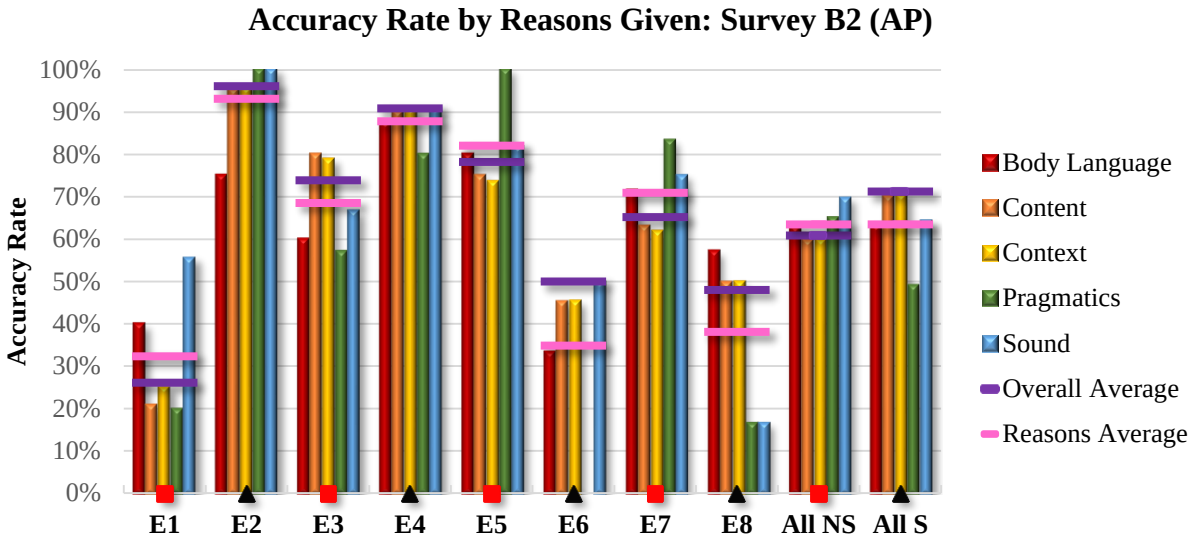

Accuracy Rate by Reasons Given: Survey B2 (AP)
Accuracy Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
E1
E2
E3
E4
E5
E6
E7
E8
All NS
All S
Body Language
Content
Context
Pragmatics
Sound
Overall Average
Reasons Average

Figure 4.41.1.

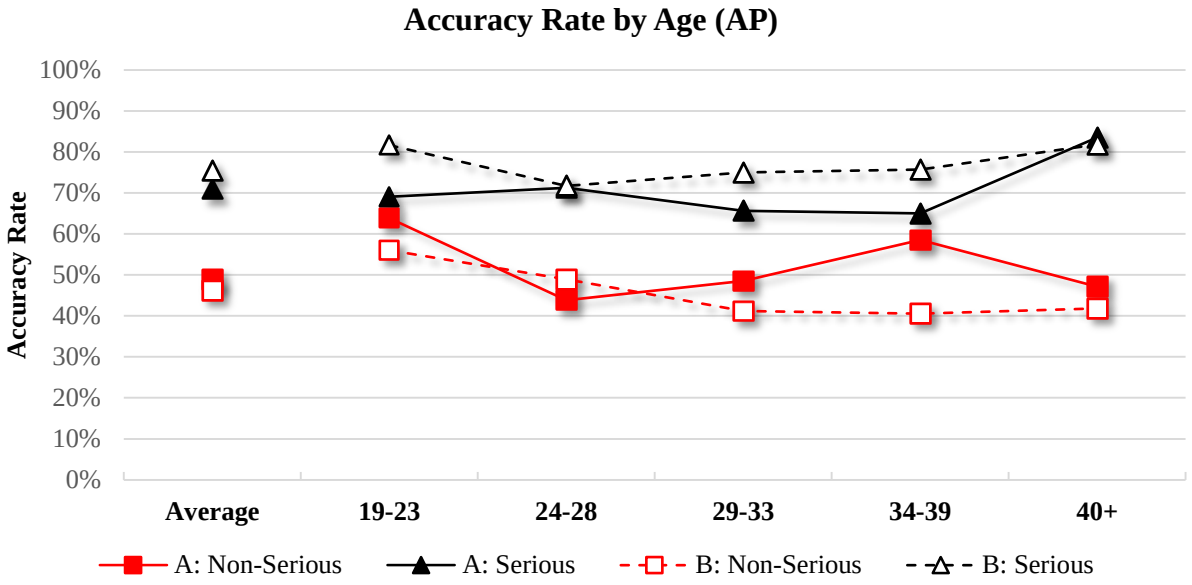

Accuracy Rate by Age (AP)
Accuracy Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Average
19-23
24-28
29-33
34-39
40+
A: Non-Serious
A: Serious
B: Non-Serious
B: Serious

Figure 4.42.1

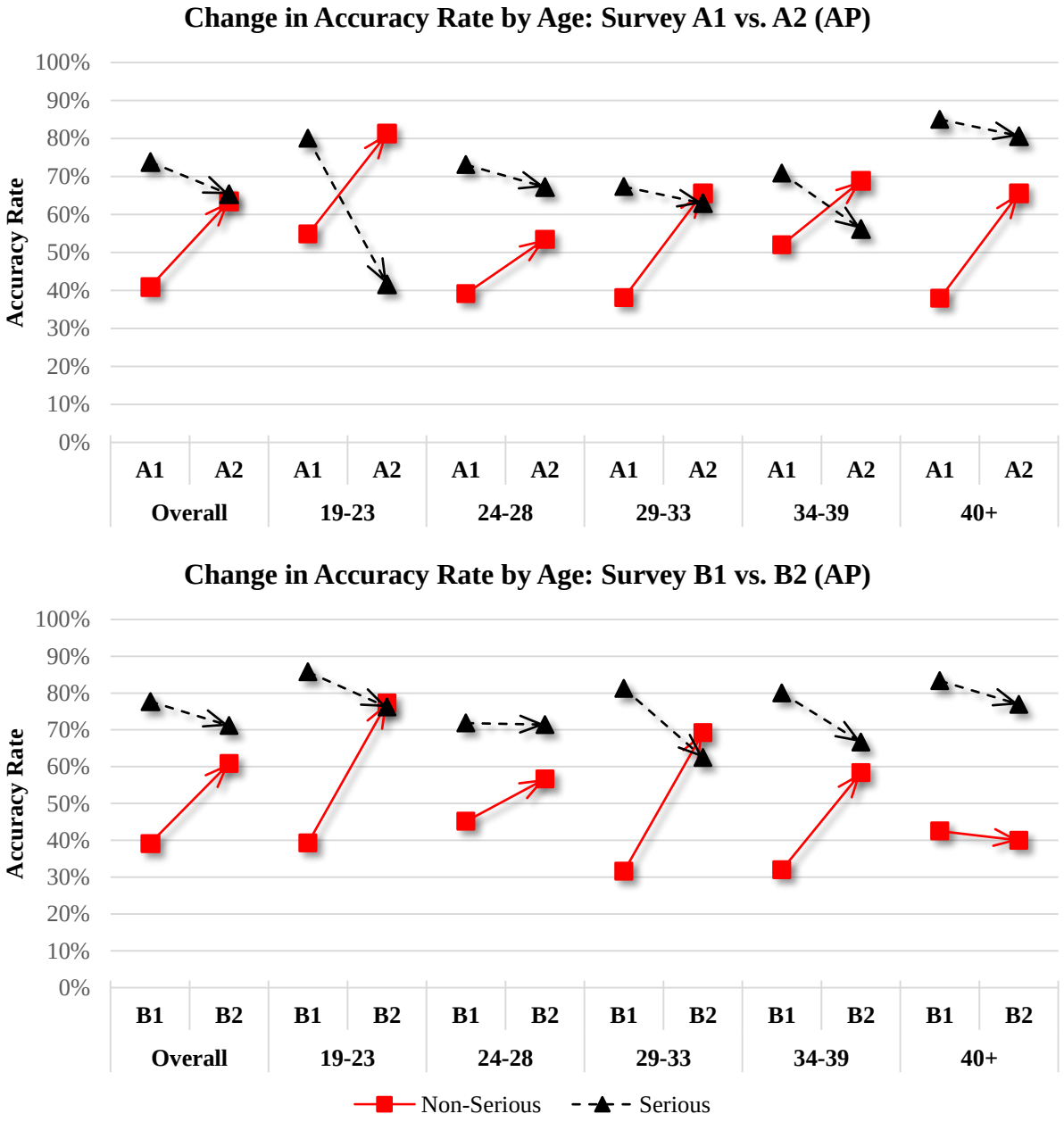

Change in Accuracy Rate by Age: Survey A1 vs. A2 (AP)
Accuracy Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A1
A2
Overall
19-23
24-28
29-33
34-39
40+
Change in Accuracy Rate by Age: Survey B1 vs. B2 (AP)
B1
B2
Non-Serious
Serious

Figure 4.43.1.

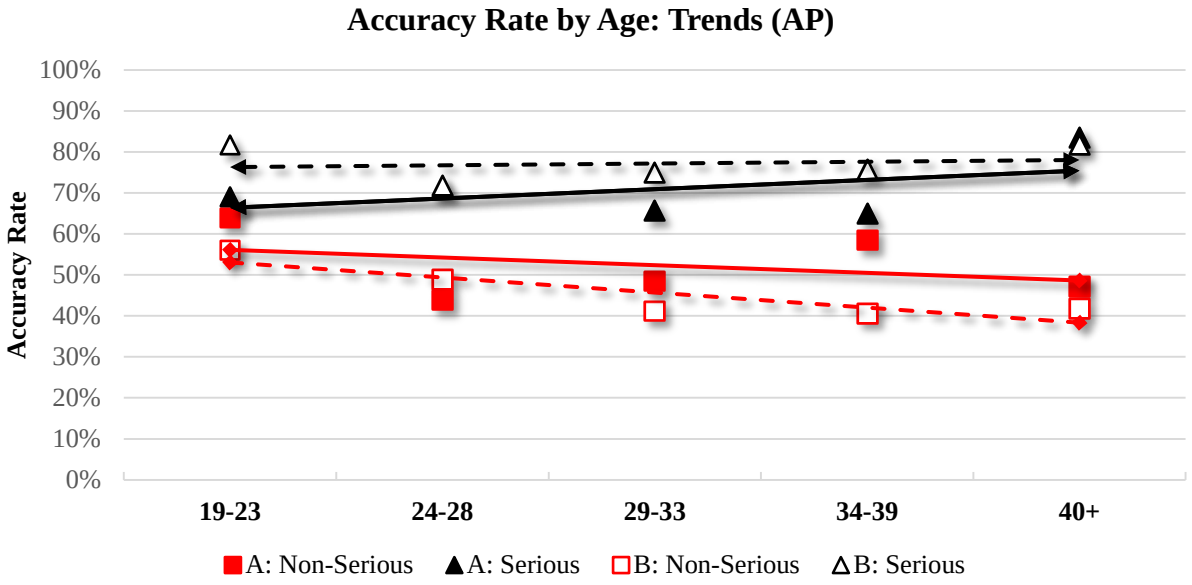

Accuracy Rate by Age: Trends (AP)
Accuracy Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
19-23
24-28
29-33
34-39
40+
A: Non-Serious
A: Serious
B: Non-Serious
B: Serious

Figure 4.44.1.

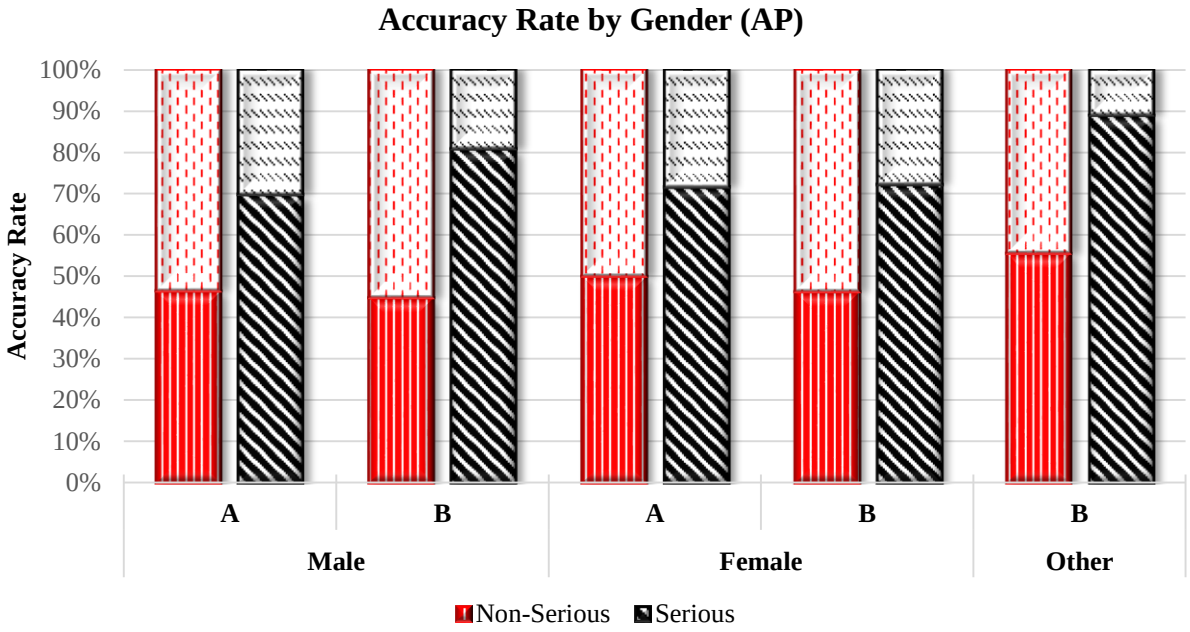


Figure 4.45.1.

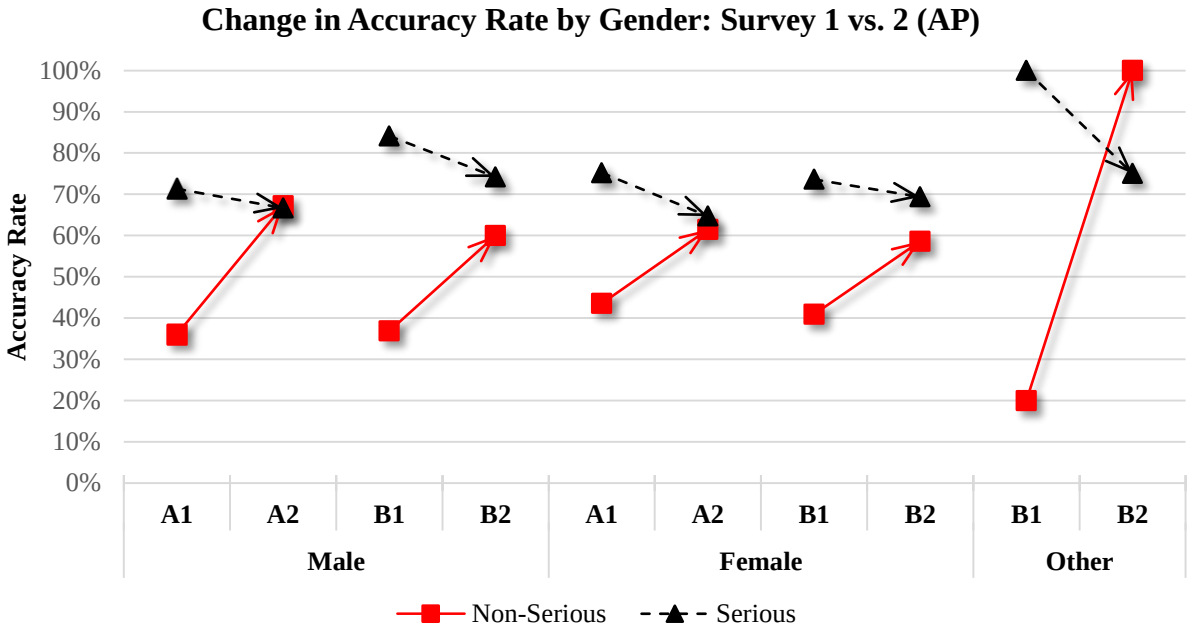


Figure 4.46.1.

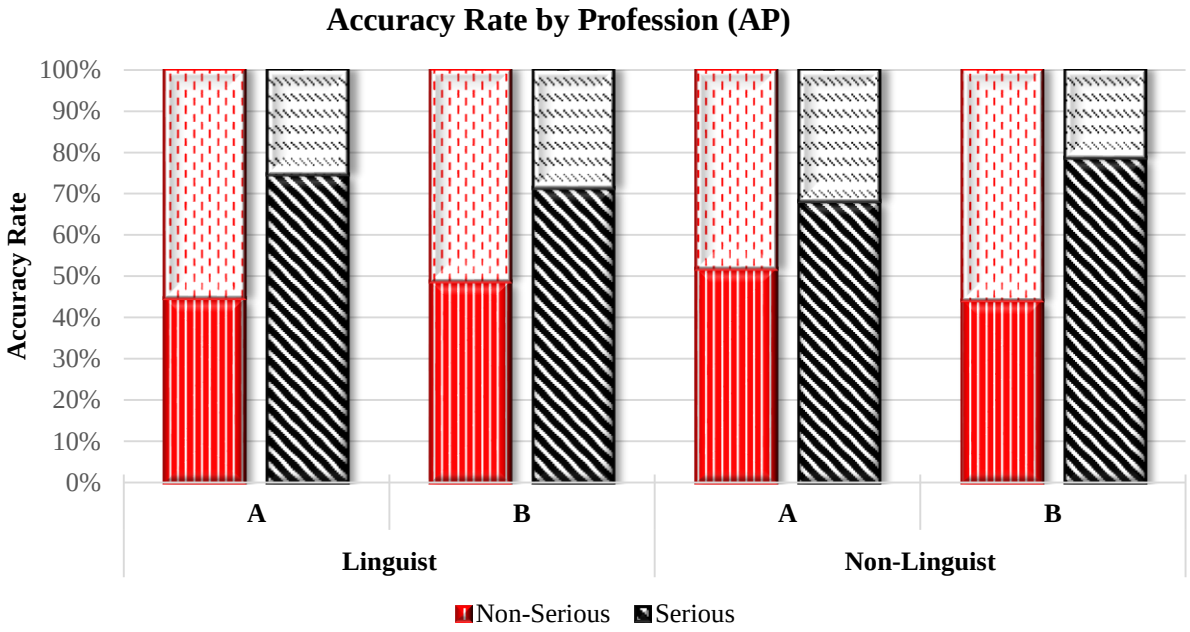


Figure 4.47.1.