# Investigations in Mechanisms and Strategies to Enhance Hearing with Cochlear Implants

By

Tyler H. Churchill

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Physics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2013

Date of final oral examination: 11/05/13

The dissertation is approved by the following members of the Final Oral Committee:

Ruth Litovsky, Professor, Communication Sciences and Disorders
Robert Lutfi, Professor, Communication Sciences and Disorders
Sue Coppersmith, Professor, Physics
Stefan Westerhoff, Professor, Physics
Francis Halzen, Professor, Physics

To all those on whose shoulders I've stood or rested.

Acknowledgements

The research performed for this dissertation was supported by the NIH-NIDCD (5R01 DC003083, Litovsky) and also in part by a core grant to the Waisman Center from the NICHD (P30 HD03352). Computing resources of the University of Wisconsin's Center for High Throughput Computing were used extensively.

Table of Contents

Dedication i

Acknowledgements ii

Table of Contents iii

Abstract iv

Chapter 1: Background and introduction 1

Chapter 2: Speech perception with a harmonic complex vocoder 17

Chapter 3: Exploring location cue encoding in the auditory periphery with a computational model 46

Chapter 4: Spatial hearing with speech temporal fine structure in bilateral cochlear implant listeners 87

Chapter 5: Discussion and conclusions 122

Abstract

Cochlear implants (CIs) produce hearing sensations by stimulating the auditory nerve (AN) with current pulses whose amplitudes are modulated by filtered acoustic temporal envelopes. While this technology has provided hearing for multitudinous CI recipients, even bilaterally-implanted listeners have more difficulty understanding speech in noise and localizing sounds than normal hearing (NH) listeners. Three studies reported here have explored ways to improve electric hearing abilities.

Vocoders are often used to simulate CIs for NH listeners. Study 1 was a psychoacoustic vocoder study examining the effects of harmonic carrier phase dispersion and simulated CI current spread on speech intelligibility in noise. Results showed that simulated current spread was detrimental to speech understanding and that speech vocoded with carriers whose components' starting phases were equal was the least intelligible. Cross-correlogram analyses of AN model simulations confirmed that carrier component phase dispersion resulted in better neural envelope representation.

Localization abilities rely on binaural processing mechanisms in the brainstem and mid-brain that are not fully understood. In Study 2, several potential mechanisms were evaluated based on the ability of metrics extracted from stereo AN simulations to predict azimuthal locations. Results suggest that unique across-frequency patterns of binaural cross-correlation may provide a strong cue set for lateralization and that interaural level differences alone cannot explain NH sensitivity to lateral position.

While it is known that many bilateral CI users are sensitive to interaural time differences (ITDs) in low-rate pulsatile stimulation, most contemporary CI processing strategies use high-rate, constant-rate pulse trains. In Study 3, we examined the effects of pulse rate and pulse timing on ITD discrimination, ITD lateralization, and speech recognition by bilateral CI listeners. Results showed that listeners were able to use low-rate pulse timing cues presented redundantly on multiple electrodes for

ITD discrimination and lateralization of speech stimuli even when mixed with high rates on other electrodes.

These results have contributed to a better understanding of those aspects of the auditory system that support speech understanding and binaural hearing, suggested vocoder parameters that may simulate aspects of electric hearing, and shown that redundant, low-rate pulse timing supports improved spatial hearing for bilateral CI listeners.

# Chapter 1

## Background and Introduction

Hearing loss is a ubiquitous medical condition, affecting hundreds of thousands of Americans and millions worldwide, and burdens significant costs upon society and the individuals afflicted (Mohr et al. 2000; Nachtegaal et al. 2012). Hearing loss can be broadly divided between two non-exclusive categories: conductive and sensorineural. Conductive hearing loss describes the condition wherein sound is not effectively conducted through the outer ear and middle ear structures to the inner ear transduction organ, the cochlea. This may be due to obstruction, malformation of the outer or middle ear structures, restriction of middle ear structure motion by fluids, or otosclerosis. If correctly identified, treatments attempt to improve the sound conduction pathway or to provide for an alternative conduction pathway to the cochlea, and may involve surgery or the implantation of an auditory prosthesis. If such treatments are unsuccessful or unavailable, or if the hearing loss is improperly

diagnosed, hearing aids may be prescribed. In the case of improper diagnosis, the resulting exposure to elevated sound levels may also lead to sensorineural hearing loss (SNHL).

Sensorineural hearing loss describes the condition wherein the cochlear transduction mechanism or higher nervous system functions are inhibited. Sensorineural hearing loss arises from myriad sources. It may be congenital, grow over time, or occur due to a singular damaging event. Contributing non-genetic causes include diseases such as rubella, measles, meningitis, and otitis media; exposure to ototoxic pharmacological agents; aging (presbycusis); and prolonged and/or acute exposure to excessive noise (Tucci et al. 2010). Although preventative measures such as immunization and the use of hearing protection are highly effective for some etiologies, corrective measures are necessary for a large portion of the afflicted population. The primary treatment for mild to moderate SNHL is the hearing aid (HA). However, HA use is a one-way street with SNHL. The elevated sound levels which provide better hearing for HA users also contribute to further damage to the cochlea.

The machinery for the normal transduction of sound into neural signals is both elegant and fragile, and deserves a brief summary. In normal hearing, sound pressure waves enter through and are filtered by the outer ear and impinge upon the tympanic membrane, the eardrum. The chain of middle ear bones, the ossicles, act as a lever which couples the motion of the tympanic membrane to that of the membrane which covers the oval window on the base of the cochlea. Thus the middle ear acts as an impedance-matching coupler between sound's natural medium, air, and the cochlear fluid. Positive pressure exerted upon the oval window is conveyed hydraulically to and relieved by the outward motion of the second of the two cochlear windows, the round window. The spiral-shaped cochlea is partitioned longitudinally by the basilar membrane (BM) for all but a small gap at its apex, called the helicotrema. Both the oval and round windows are located on the base of the cochlea, but are separated inside the

cochlea by the BM. Hence hydraulic pressure must travel through or around (at the helicotrema) the BM. For hydrodynamic waves to travel through the BM, they must be of frequency at or near the resonance frequency of the BM. The mass and stiffness of the BM vary monotonically from the base of the cochlea to the apex. Its linear mass density increases and its stiffness decreases from base to apex, resulting in a monotonic function of the BM's resonance frequency along its length. Therefore, high frequencies pass through the BM near the base of the cochlea, while low frequencies pass through the BM near the apex of the cochlea, and each location along the length of the cochlea is associated with a characteristic frequency (CF) (Békésy 1963; Greenwood 1990). The transduction of sound occurs due to the resonant motion of the BM as hydrodynamic waves pass through it.

On the BM and along its length sits the Organ of Corti (OC), which contains three rows of outer hair cells (OHCs) and one row of inner hair cells (IHCs), and atop each of the tens of thousands of these hair cells is a bundle of finger-like stereocilia. Hovering directly above the BM, OC, hair cells, and their stereocilia is the tectorial membrane (TM). Transverse motion of the BM results in relative lateral motion between the OC (with its imbedded hair cells and their stereocilia) and the TM. Through direct or fluid coupling, this differential motion causes the affected IHCs' stereocilia to sway and their tip-linked mechanotransduction ion channels to open, depolarizing the IHCs. Due to the anatomical geometry of hair cell bundles, IHCs tend to depolarize at a certain phase of the BM motion. The precise phase is dependent upon cochlear location and maximum BM motion amplitude, but the general phenomenon is known as "phase locking" (Ruggero and Rich 1987). Upon IHC depolarization, neurotransmitters may be released into synapses connecting the IHC to an auditory nerve (AN), possibly resulting in the generation of an action potential in the AN. (The use of the terms "may be" and "possibly" reflect the fact that neurotransmitter release and detection are stochastic processes, and in fact produce spontaneous activity in auditory neurons.) For frequencies below several kHz, signals on

the AN can thus be phase-locked to the waveform oscillation patterns (Javel and Mott 1988). The AN conducts the neural signals to the cochlear nucleus (CN) in the brainstem, from whence they are transmitted to other nerve centers in the brain which are responsible for the percepts of hearing. Damage to any of these structures may contribute to SNHL, but it is thought that broken stereocilia tip links may be a common factor.

The peripheral afferent auditory system is accompanied by a parallel efferent system. Auditory nerves also synapse onto the OHCs, which expand and contract in response to signals from the AN. The motion of the OHCs, whose stereocilia may be physically connected to the TM, influences the relative motion of the BM and the TM such that the region of effective resonance along the BM's length is narrowed and transduction at the IHCs is enhanced near the resonance point, resulting in sharper frequency tuning. Hence, OHC function is of paramount importance for the transduction of unique frequencies and also for hearing faint sounds. Damage to OHCs may precurse IHC damage on the road to SNHL.

Having two normally-functioning ears not only provides valuable redundancy in the sensory system (Viemeister and Wakefield 1991), but also facilitates many aspects of spatial hearing. The head acts to shadow one ear from a sound coming from the contralateral side, producing an interaural level difference (ILD) between the two ears. Additionally, the difference in arrival time between the two ears for a laterally-positioned sound source results in an interaural time difference (ITD). The binaural processing centers of the auditory system produce a perception of azimuthal location based on the left/right differential encoding of these physical cues by the auditory periphery. In addition to sound localization, binaural hearing contributes to better hearing in complex environments by suppressing echoes, a phenomenon known as the "precedence effect" (Litovsky et al. 1999) and also by allowing

listeners to attend to spatially unique sources, introducing auditory stream segregation as a solution to the phenomenon known as the “cocktail party problem” (Cherry 1953; Bronkhorst 2000).

That the application of electric currents to the ear can induce heard sensations has been known for over two hundred years (Volta 1800), but it was not ascertained until the mid-20th century that electrical stimulation of the AN was responsible for these sensations (Jones et al. 1940). In 1965, results were published of an experiment in which several electrodes were implanted into the auditory nerve of a deaf person, near the base of the cochlea (Simmons et al. 1965). Electrical stimulation consisting of biphasic pulse trains or sinusoids was applied, and reported sensations recorded. Psychoacoustic testing concluded that perceived loudness increased with applied current levels and that pitch was dependent upon both place and rate of stimulation. Further work by Michaelson (1971) led to the introduction of the first single-electrode cochlear implant (CI) auditory prosthesis by House (1976). These early CIs consisted of little more than a microphone connected to a processor which controlled a current source for the implanted electrode. Perhaps they mere provided an indication of the presence of sound and lip reading cues for profoundly deaf persons, but their contributions to contemporary and future CI users’ quality of life is immeasurable. The technology has advanced, and present-day CIs use multiple-electrode arrays and can provide users excellent open-set speech understanding in quiet without lip reading (Moller 2006).

Contemporary CIs consist of two primary components- the implanted device and the external, worn device. The implanted part includes the array of electrodes that is implanted into the cochlea, reference/ground electrodes, a magnet and coil for the transcutaneous RF communication link, and a hermetically-sealed processor that houses the power source, decoder, and current source (Zeng et al. 2008). There are no internal batteries- the device is powered by RF it receives from the external device.

Besides the addition of electrodes, little has changed over the years regarding the internal device. The external device, on the other hand, has evolved from a large, body-worn processor to a small, behind-the-ear processor which is held in place by the implanted magnet. The external device consists of a microphone, signal processing hardware, and an RF stimulator, and is powered by disposable or rechargeable batteries. Although there are configuration differences among the different manufacturers' devices, the general architecture is the same.

Signal processing for CIs has also greatly changed over the years. The overarching goal of CI signal processing is to transform acoustic information that is relevant to speech understanding into electrical signals which will activate the peripheral auditory system in such a way as to adequately convey said acoustic information. Due to the tonotopic organization of the cochlea, by electrically stimulating the AN using electrodes at different cochlear insertion depths, different frequency percepts are elicited. Therefore, speech signals that are broken-down by a set of band-pass filters into a number of channels can be presented to the auditory periphery by reproducing channel-specific speech information via electrical stimulation at the corresponding electrodes. Hence, CIs present speech information from high-frequency channels through electrical stimulation on basally-implanted electrodes and information from low-frequency channels on apically-implanted electrodes. Louder signals are represented by higher levels of current, which in turn evoke louder percepts.

Electrical stimulation with early CIs was in the form of an AC analog signal that was a compressed version of the acoustic waveform at the output of each channel's band-pass filter (Zeng 2004). That method is the most straightforward transformation of multi-channel acoustic to multi-channel electric information, but the peripheral auditory system does not respond to electrical AC waveforms the same way it responds to acoustic pressure waves. For example, the firing patterns on the

AN lose much of their stochasticity when exposed to electric stimulation and exhibit “super” phase locking to the AC electric waveform (Kiang and Moxon 1972). The foremost problem for speech understanding seemed to be the interaction of simultaneously active electrodes’ electric fields (Loizou 1997). In order to avoid these channel interaction effects due to electric field overlap with simultaneous stimulation on multiple electrodes, continuous analog signals were replaced by channel-interleaved, amplitude-modulated, biphasic, square current pulse trains. The first CI processing strategy to adopt these methods was the aptly-named Continuous Interleaved Sampling (CIS) strategy (Wilson et al. 1991). With CIS and other pulsatile strategies, the individual pulses’ current magnitudes are modulated by their respective channels’ peak amplitude temporal envelopes. These methods remove the interaction problems of simultaneous stimulation, but since most pulsatile strategies use constant pulse rates, much potentially-useful waveform information is also discarded.

The most widely-used clinical CI processing strategy is the Advanced Combination Encoder (ACE) (Arndt et al. 1999). Unlike CIS, which stimulates on all available electrodes during a several-millisecond interleaving cycle, ACE chooses “n of m” electrodes on which to stimulate, where m is the number of available electrodes, and n is the number of electrodes on which to stimulate in a given cycle, with $n<m$. The n electrodes are chosen as those which correspond to channels containing the n highest amounts of energy on a given cycle. Thus, spectral peaks are selectively stimulated while low-energy channels are ignored. The ACE strategy has been found to produce enhanced speech recognition, among other perceptual improvements (Kiefer et al. 2001). Nearly all contemporary strategies, including ACE and CIS, use high stimulation rates (>900 pulses per second on each electrode).

Cochlear implants have been remarkably successful, with most adult listeners achieving excellent open-set speech recognition with a year of activation (Wilson and Dorman 2008; Lenarz et al.

2012; Gaylor et al. 2013) and children achieving significant improvements in spoken language scores (Niparko et al. 2010). This success has resulted in expanding candidate criteria (Lusis 2005; Wiley and Meinzen-Derr 2009). Bilateral implantation is implemented in order to attempt to reproduce the benefits of binaural hearing, and has been shown to improve spatial hearing for children (Litovsky et al. 2006) and adults (Litovsky et al. 2010), and to further improve speech perception in adults, relative to with unilateral implants alone (Gaylor et al. 2013). However, speech understanding with CIs is still significantly worse than with normal hearing (NH), especially in noise (Friesen et al. 2001), and the gains in spatial hearing due to bilateral implantation have been less than would be expected with restored binaural hearing. Many CI users exhibit spatial hearing abilities that appear to be informed solely by head shadow or ILD effects (van Hoesel et al. 2008), and their ITD sensitivity is both considerably lower than with normal binaural hearing and limited to pulse rates below several hundred Hz (van Hoesel et al. 2009). Because CIs are only designed to work in a single ear, the trend toward bilateral implantations has positioned the field in a challenging situation of having to decide the importance of providing faithful binaural cues versus simply bilateral cues.

Although non-technological factors such as neural survival, plasticity and learning, and central mechanisms no doubt also contribute to individuals' outcomes with CIs, the performance gap between CI and NH listeners has been at least partially attributed to deficits in both spectral and temporal resolution with electric hearing when compared to NH mechanisms (Moore 2003). Speech recognition performance as a function of the number of electrodes used asymptotes above 7, indicating that even when more stimulating electrodes are available, the spread of current activates sufficient swaths of auditory neurons so as to render stimulation on nearby electrodes, if not indistinguishable, at best ineffective for the conveyance of unique information for improved speech recognition. The inability to exert precise control over the path of current (and resulting AN activation) from the active electrode to

the reference ground using biphasic square pulses thus limits the spectral resolution available. Several researchers are currently exploring the consequences of bi-, tri-, and multi-polar stimulation in order to investigate possible solutions to the problem of poor spectral resolution due to current spread (Berenstein et al. 2008).

Similarly, much effort has been devoted to exploring the benefits of presenting more temporal information regarding the acoustic waveforms via the timing of CIs' electrical pulses. The majority of contemporary processing strategies on commercially-available CIs use constant-rate pulse train stimulation that ignores waveform subtleties. Since the slowly-varying temporal envelopes are themselves sufficient for understanding multichannel speech signals in quiet (Shannon et al. 1995), ignoring these envelopes' acoustic carriers has been the common, and an adequate, approach for CIs. However, the temporal information contained in the waveforms carrying these envelopes, the temporal fine structure (TFS), which is naturally encoded by phase-locking in normal hearing, is known to be important for understanding speech in noise and sound localization (Drennan et al. 2007). One exception, manufacturer MED-EL's Fine Structure Processing strategy (FSP), uses Channel-Specific Sampling Sequences, which times the electric pulses on a given electrode to represent the acoustic TFS in the corresponding channel (Zierhofer 2003). Although super phase locking might be expected to give FSP's users super temporal resolution, this approach has not met with spectacular clinical success (Magnusson 2011), possibly due to current spread (Colburn et al. 2009).

Thus many challenges yet remain in improving the world's most successful brain-computer interface. The uncertainties regarding the path of electrical current through the cochlea and the interaction of information on AN fibers of different CF at more central processing nuclei leaves us with questions of how to best tackle these challenges and a wealth of possible research directions. However,

this wealth is balanced by a poverty in the number of bilaterally-implanted CI users who are available for testing. Although we are rightly limited in the configurations we can impose on and the recordings we can make from human CI users, we have a substantial set of parameters to vary and psychophysical protocols to employ. In addition, we can simulate the experience of electric hearing in NH listeners using a device called a vocoder, and greatly expand our subject pool.

The channel vocoder is a tool, normally implemented in software, which filters input speech signals into contiguous frequency pass-bands, calculates the temporal envelope at the output of each one of these bands, modulates band-limited carriers by these envelopes, and recombines the resulting signals. The intelligibility of the output speech depends on the vocoder's parameters, i.e., the carrier chosen, the filter specifications, and the number of channels used. Variations on the original vocoder (Dudley 1939; Schroeder 1966) have been used extensively in psychoacoustic experiments to study aspects of speech recognition, and CI signal processing relies on the same type of front-end analyses as the channel vocoder (Loizou 2006). Thus, it is natural to attempt to simulate electric hearing using a vocoder and therefore imperative to assess which choices of parameters result in the best simulation of CI hearing (Dorman et al. 1997; Qin and Oxenham 2003; Whitmal et al. 2007). Much literature has tacitly assumed the equivalence of electric hearing and vocoder hearing, but we must be careful to justify our assumptions about the connections among NH, vocoder listening, and CI listening.

This dissertation consists of reports from three studies that have explored NH and CI mechanisms and strategies to enhance hearing with CIs. The goal of these studies was to better understand those aspects of the peripheral auditory system that support speech understanding and binaural hearing, to recommend vocoder parameters that simulate certain aspects of electric hearing

based on an exploration of that parameter space, and to propose and test a novel CI processing strategy for improved spatial hearing.

Chapter 2 reports the results of a vocoder experiment with NH subjects. The experiment assessed speech recognition in noise as a function of signal-to-noise ratio, carrier type, and synthesis filter shape, and tested the hypothesis that randomizing the relative starting phases of a complex harmonic carrier would result in better speech recognition. The study also examined the neural representation of this vocoded speech at the auditory periphery using a computational AN model and assessed the contribution of various cues to speech understanding using shuffled cross-correlogram analyses. The results suggest which carrier temporal and spectral characteristics best simulate CI listening and which characteristics produce better speech understanding.

Chapter 3 reports an extension of the analyses used in Chapter 2 to binaural hearing and an exploration of the mechanisms of binaural processing in the brainstem. The experiment calculated and evaluated several metrics from the outputs of the computation AN model for stereo stimuli with various azimuth and cue configurations. These metrics were analyzed for their uniqueness to the generating azimuth and for their ability to predict psychoacoustic localization response data. The results suggest that mechanisms associated with several of the different metrics may contribute to perceived lateral position, and that the products of these mechanisms may be perceptually weighted by their consistency.

Chapter 4 reports the results of a direct stimulation experiment with bilaterally-implanted CI listeners. The experiment assessed speech recognition in quiet, ITD lateralization, and ITD discrimination as functions of strategy and stimulation pulse rate used, and tested the hypothesis that TFS-timed pulses presented at low rates, redundantly across multiple channels, would produce significant improvements in ITD sensitivity. The experiment also tested a subset of listeners on speech recognition

in noise and calculated speech reception thresholds as functions of strategy and stimulation pulse rate. The results show that TFS-timed pulses result in improved speech recognition and better ITD sensitivity, and quantify a trade-off between speech understanding and ITD sensitivity as a function of stimulating pulse rates used.

**References**

Arndt P, Staller S, Arcaroli J, et al. (1999) Within-subject comparison of advanced coding strategies in the Nucleus 24 cochlear implant (Cochlear Corporation Report for FDA submission).

Békésy, G. Von. (1963). Hearing theories and complex sounds. The Journal of the Acoustical Society of America, 35(4), 588–601.

Berenstein CK, Mens LHM, Mulder JJS, Vanpoucke FJ (2008) Current steering and current focusing in cochlear implants: comparison of monopolar, tripolar, and virtual channel electrode configurations. Ear and hearing 29:250–60.

Bronkhorst AW (2000) The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. Acustica 86:117–128.

Cherry E (1953) Some experiments on the recognition of speech with one and two ears. J Acoust Soc Am 25:975–979.

Colburn HS, Chung Y, Zhou Y, Brughera A (2009) Models of brainstem responses to bilateral electrical stimulation. Journal of the Association for Research in Otolaryngology : JARO 10:91–110. doi: 10.1007/s10162-008-0141-z

Dorman MF, Loizou PC, Rainey D (1997) Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. The Journal of the Acoustical Society of America 102:2403–2411.

Drennan WR, Won JH, Dasika VK, Rubinstein JT (2007) Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise. Journal of the Association for Research in Otolaryngology : JARO 8:373–83. doi: 10.1007/s10162-007-0074-y

Dudley H (1939) Remaking Speech. J Acoust Soc Am 11:169–177.

Friesen LM, Shannon R V, Baskent D, Wang X (2001) Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. The Journal of the Acoustical Society of America 110:1150. doi: 10.1121/1.1381538

Gaylor J, Raman G, Chung M, et al. (2013) Cochlear Implantation in Adults. JAMA Otolaryngol Head Neck Surg 139:265–272. doi: 10.1001/jamaoto.2013.1744

Greenwood DD (1990) A cochlear frequency-position function for several years later. 2592–2605.

House W (1976) Cochlear Implants. Ann Otol Rhinol Laryngol 85:1–93.

Javel E, Mott JB (1988) Physiological and psychophysical correlates of temporal processes in hearing. Hearing research 34:275–294.

Jones R, Stevens S, Lurie M (1940) Three mechanisms of hearing by electrical stimulation. The Journal of the Acoustical Society … 12:261–269.

Kiang NY, Moxon EC (1972) Physiological considerations in artificial stimulation of the inner ear. Ann Otol Rhinol Laryngol 81:714–730.

Kiefer J, Hohl S, Sturzebecher E, et al. (2001) Speech Recognition with Different Speech Coding Strategies (SPEAK, CIS, and ACE) and Their Relationship to Telemetric Measures of Compound Action Potentials. Audiology 40:32–42.

Lenarz M, Sönmez H, Joseph G, et al. (2012) Long-term performance of cochlear implants in postlingually deafened adults. Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery 147:112–8. doi: 10.1177/0194599812438041

Litovsky RY, Colburn HS, Yost W a, Guzman SJ (1999) The precedence effect. The Journal of the Acoustical Society of America 106:1633–54.

Litovsky RY, Johnstone PM, Godar SP (2006) Beneftis of bilateral CIs and/or HAs in children. International journal of audiology 45:1–22. doi: 10.1080/14992020600782956.Benefits

Litovsky RY, Parkinson A, Arcaroli J (2010) Spatial Hearing and Speech Intilligibility in Bilateral Cochlear Implant Users. Ear and hearing 30:419–431. doi: 10.1097/AUD.0b013e3181a165be.Spatial

Loizou PC (1997) Signal Processing for Cochlear Prosthesis : A Tutorial Review. IEEE

Loizou PC (2006) Speech processing in vocoder-centric cochlear implants. Advances in oto-rhino-laryngology 64:109–43. doi: 10.1159/000094648

Lusis I (2005) Access Expands for Cochlear Implants : CMS Decision Affects Eligibility Criteria. The ASHA Leader

Magnusson L (2011) Comparison of the fine structure processing (FSP) strategy and the CIS strategy used in the MED-EL cochlear implant system: speech intelligibility and music sound quality. International journal of audiology 50:279–87. doi: 10.3109/14992027.2010.537378

Michaelson R (1971) Electrical Stimulation of the Human Cochlea: A preliminary report. Arch Otolaryngol 93:317–323.

Mohr PE, Feldman JJ, Dunbar JL (2000) The societal costs of severe to profound hearing loss in the United States. Policy analysis brief H series / Project Hope, Center for Health Affairs 2:1–4.

Moller A (2006) A History of Cochlear Implants and Auditory Brainstem Implants. In: Moller A (ed) Cochlear and Brainstem Implants. pp 1–10

Moore BCJ (2003) Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology 24:243–54.

Nachtegaal J, Festen JM, Kramer SE (2012) Hearing ability in working life and its relationship with sick leave and self-reported work productivity. Ear and hearing 33:94–103. doi: 10.1097/AUD.0b013e318228033e

Niparko J, Tobey E, Thal D, et al. (2010) Spoken language development in children following cochlear implantation. Journal of the American Medical Association 303:1498–1506.

Qin MK, Oxenham AJ (2003) Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. The Journal of the Acoustical Society of America 114:446. doi: 10.1121/1.1579009

Ruggero MA, Rich NC (1987) Timing of spike initiation in cochlear afferents: dependence on site of innervation. Journal of neurophysiology 58:379–403.

Schroeder MR (1966) Vocoders: Analysis and Synthesis. Proceedings of the IEEE 54:720–734.

Shannon R V., Zeng F-G, Kamath V, et al. (1995) Speech Recognition with Primarily Temporal Cues. Science 270:303–304.

Simmons F, Epley J, Lummis R, et al. (1965) Auditory nerve: electrical stimulation in man. Science 148:104–106.

Tucci D, Merson MH, Wilson BS (2010) A summary of the literature on global hearing impairment: current status and priorities for action. Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology 31:31–41.

Van Hoesel RJM, Böhm M, Vandali A, et al. (2008) Binaural speech unmasking and localization in noise with bilateral cochlear implants using envelope and fine-timing. J Acoust Soc Am 12:2249–2263. doi: 10.1121/1.2875229

Van Hoesel RJM, Jones GL, Litovsky RY (2009) Interaural time-delay sensitivity in bilateral cochlear implant users: effects of pulse rate, modulation rate, and place of stimulation. Journal of the Association for Research in Otolaryngology : JARO 10:557–567. doi: 10.1007/s10162-009-0175-x

Viemeister NF, Wakefield GH (1991) Temporal integration and multiple looks. The Journal of the Acoustical Society of America 90:858–865.

Volta A (1800) On the electricity excited by the mere contact of conducting substances of different kinds. Philosophical transactions of the Royal Society of London 90:403–431.

Whitmal NA, Poissant SF, Freyman RL, Helfer KS (2007) Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. The Journal of the Acoustical Society of America 122:2376–88. doi: 10.1121/1.2773993

Wiley S, Meinzen-Derr J (2009) Access to cochlear implant candidacy evaluations: who is not making it to the team evaluations? International journal of audiology 48:74–9. doi: 10.1080/14992020802475227

Wilson B, Finley C, Lawson D, et al. (1991) Better speech recognition with cochlear implants. Nature 352:236–238.

Wilson BS, Dorman MF (2008) Cochlear implants: a remarkable past and a brilliant future. Hearing research 242:3–21. doi: 10.1016/j.heares.2008.06.005

Zeng FG, Rebscher S, Harrison W, Sun X ; Feng H (2008) Cochlear Implants : System Design , Integration , and Evaluation. 1:115–142.

Zeng FG (2004) Trends in cochlear implants. Trends in amplification 8:1–34.

Zierhofer CM (2003) Electrical nerve stimulation based on channel specific sampling sequences.

# Chapter 2

# Speech perception with a harmonic complex vocoder

**Introduction**

The ability to recognize speech in noise with cochlear implants (CIs) has not yet achieved parity with normal hearing (NH) (Friesen et al. 2001; Van Deun et al. 2010; Eskridge et al. 2012), likely a result of the poorer spectral and temporal resolution in electric hearing (Fu et al. 2004; Fu and Nogaki 2004). The channel vocoder has been used extensively to study aspects of speech understanding (Dudley 1939; Schroeder 1966; Shannon et al. 1995). Because its principles are employed in CI processing (Loizou 2006), the vocoder has often been used to simulate electric hearing for NH listeners (Dorman et al. 1997; Fu et al. 1998; Nelson et al. 2003; Chen and Loizou 2011). By comparing how NH and CI listeners understand degraded speech signals, the vocoder parameters that best simulate electric hearing and factors that might contribute to the known gap in performance between NH and CI listeners are better understood. Like CIs, vocoders discard acoustic temporal fine structure (TFS) and present only band-

limited envelope information from the original waveform. These envelopes modulate the vocoder's carrier. Although simple tones and filtered noise are the most commonly-used carriers, Gaussian-enveloped tones (Lu et al. 2007) and harmonic complexes have also been used (Deeks and Carlyon 2004; Hervais-Adelman et al. 2011). It is unclear which vocoder carrier best approximates electric hearing, but different carriers result in different speech intelligibility depending on the parameters chosen. Whitmal et al. (2007) found that sine vocoders, with flat intrinsic carrier envelopes and prominent sidebands, resulted in better modulation detection and speech understanding than noise vocoders. Using a lower envelope cutoff frequency than that study, Hervais-Adelman et al. (2011) found that sine-vocoded speech was more difficult to understand than noise-vocoded speech, for a fixed modulation depth. These studies demonstrate that the carrier characteristics are important parameters affecting speech understanding with vocoders.

The present study tested speech recognition in noise for several different carriers. In addition, stimuli were generated using synthesis filters that simulated channel interaction caused by the spread of current in CI stimulation. The hypothesis that randomly dispersing the component starting phases of a harmonic complex carrier should flatten the carrier's intrinsic envelopes and improve speech recognition was tested. Sine tone and noise carriers were tested in addition to the harmonic complex carriers as control conditions.

Fidelity of neural encoding of envelope and TFS information as measured by neural cross-correlation coefficients has previously been shown to predict perceptual identification scores for vocoded speech in noise (Swaminathan and Heinz 2012). However, the assertion that the auditory system independently encodes envelope and TFS may be suspect (Shamma and Lorenzi 2013). Here, responses of an auditory nerve (AN) model (Zilany et al. 2009) to the vocoded and unprocessed stimuli

were compared using shuffled cross-correlogram analyses (Joris 2003). Resulting neurometrics predicted scores due to the stimuli with harmonic complex carriers, but failed to predict psychoacoustic scores for all conditions, indicating that this analysis may not disentangle the differential effects of TFS and envelopes on vocoded speech intelligibility, and that more central mechanisms may play a larger role in information extraction, in agreement with Shamma and Lorenzi (2013).

| $f_{lower}$ | $f_{center}$ | $f_{upper}$ |
|---|---|---|
| 202 | 281 | 359 |
| 359 | 473 | 587 |
| 587 | 752 | 917 |
| 917 | 1156 | 1395 |
| 1395 | 1743 | 2090 |
| 2090 | 2593 | 3097 |
| 3097 | 3827 | 4558 |
| 4558 | 5617 | 6677 |

Table 1. Filter corner and center frequencies (Hz).

**Methods A: Psychoacoustics**

**Stimuli**

Fifty single-syllable, consonant-nucleus-consonant (CNC) words were vocoded using an eight-channel vocoder. Six vocoder carriers (sine tones, four different types of harmonic complexes with distinct starting phase distributions, and noise) were used to study the effects of carrier phase dispersion on speech understanding. The channel corner and center frequencies, calculated using Greenwood's function (1990) to simulate equal spacing on the cochlea, are presented in Table 1.

In order to explore the detrimental effect of simulated current spread, two sets of stimuli with different synthesis filters were generated and tested. Butterworth synthesis filters were used as a control for previously published data (Fu and Nogaki 2004) and simulated CI current spread synthesis filters (Bingabr et al. 2008) were also tested. Synthesis filters are used to filter the carrier into separate channels prior to and following modulation, whereas analysis filters are used to divide the original speech signal into separate spectral channels. Single-pass magnitude responses for each set of filters (Butterworth and "current spread") are plotted in Figure 1. Prior to vocoding, target words were mixed with a frozen token of ramped, steady-state, speech-shaped noise in order to produce stimuli with

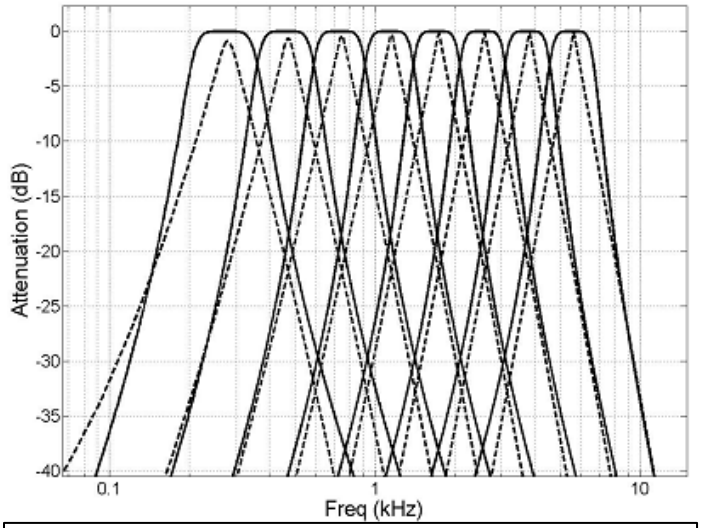



Figure 1. Butterworth (solid line) and current spread (dashed line) filters' (single-pass) magnitude responses. The Butterworth filters were used in analysis filtering for both stimuli sets and in synthesis filtering for one set of stimuli.

broadband signal-to-noise ratios (SNRs) of either 0 dB or -3 dB. The speech-shaped masker was synthesized through the inverse Fourier transform of the sum of all the CNCs' magnitude spectra with a random phase spectrum. The target was imbedded in the masker approximately one second after the masker's onset. Twelve hundred different stimuli were constructed in total (50 words × 2 SNRs × 2 synthesis filter types × 6 carrier types).

In order to vocode the stimuli, the unprocessed words were band-pass filtered into eight frequency-contiguous channels using third-order Butterworth analysis filters. Because these experiments studied the effects of phase, the filtering was performed using forward and reverse filtering, a zero phase-shift method which preserves phase relationships among components of different frequencies and results in an effective doubling of filter order. Carriers were also forward and reverse band-pass -filtered into eight channels using the appropriate synthesis filters. The signal's temporal envelopes were extracted from each channel via full-wave rectification and low-pass filtering at 50 Hz, using a second-order low-pass Butterworth filter with forward and reverse filtering. Each channel's envelope was then used to modulate the amplitude of the corresponding carrier channel. The envelope-modulated carrier channels were filtered again using their respective synthesis filters to attenuate any sidebands outside the original channel. Each vocoded stimulus was then constructed by summing its

channels. Levels for each stimulus were adjusted to ensure that all stimuli had the same root-mean-square (RMS) level. The sampling frequency was 48 kHz.

Of the two sets of synthesis filters used, the first set had identical parameters to those Butterworth filters that were used for all signals' analysis. These filters have flat magnitude responses in the pass band and overlap with the adjacent band at the half-amplitude (3-dB down) point. The second synthesis filter set was meant to better simulate the spatial dependence of current spread in CIs, and consisted of 2048-order finite impulse response (FIR) filters. These FIR filters were designed using the same center frequencies as the Butterworth filters, and were calculated to produce current decay slopes of 3.75 dB/octave. Both of these synthesis filter sets exhibit the channel overlap that is characteristic of CI current spread, but the second set, the "current spread" filters, introduce additional dynamic range compression and exhibit localized peaks with steeply decaying skirts. Adjacent Butterworth filters crossed at 3 dB of attenuation, and adjacent current spread filters crossed at 11 dB of attenuation.

Given the prevalence of sine and noise vocoders in previous studies, sine tones with frequencies equal to the channel centers and band-pass filtered white noise were used as control carriers for comparison with the harmonic complex carriers. All five carriers are summarized in Table 2. The sine and noise vocoders were denoted "S" and "N," respectively. The harmonic complex carriers were complexes of 240 equally-weighted, harmonically-spaced sine tones with a fundamental frequency of 100 Hz. Each of the harmonic complex carriers had a different component starting phase

| Carrier | Frequencies | Symbol | Phase of $n^{th}$ component |
|---|---|---|---|
| **Sine** | Filter center frequency | S | 0° |
| **Noise** | White Noise | N | N/A |
| **Harmonic complexes** | 240 equally-weighted sine harmonics with 100 Hz fundamental | H0 | 0° |
| | | H90 | random 0° - 90° |
| | | H360 | random 0° - 360° |

Table 2. Vocoder carriers consisted of two commonly-used control carriers (sine tones and white noise) and harmonic complexes with identical long-term magnitude spectra, only differing in component starting phase.

distribution. The first harmonic complex carrier was in sine phase, i.e., the starting phase of each sine tone component was zero ("H0"). Thus, it resulted in a periodic, biphasic pulse train. The second harmonic complex carrier added a random value between 0 and $\pi$ /2 to the starting phase of each component ("H90"). This processing resulted in a carrier waveform resembling a biphasic pulse train with low-amplitude noise between the pulses. The third harmonic complex carrier added a random value between 0 and $2\pi$ to the starting phase of each component ("H360"). While this carrier's time-domain signal superficially resembled noise, it elicited a 100 Hz pitch percept due to the signal's periodicity. A fourth harmonic complex carrier based on the Schroeder-minus chirp (Schroeder 1970) was constructed and tested, but due to vocoder filtering, the desired temporal characteristics were lost. The resulting waveform and performance results were nearly identical to those of the H0 stimuli, and are not discussed further. Time domain plots and spectrograms of the unprocessed CNC "goose" in quiet and three vocoded tokens thereof are shown in the top two rows of Figure 2. These illustrations allow for qualitative feature comparison among vocoder outputs. Because of the similar appearances of stimuli for H0 and H90 carriers, the latter is not shown. Carrier N, whose output has a similar appearance as that for the carrier H360, is also not shown. The plot for the original (unprocessed) CNC shows a smooth envelope and regular fine structure corresponding to the fundamental frequency ($f_0$) of voicing. Note the varying levels of broadband envelope similarity to the unprocessed signal among the vocoded samples as illustrated in the plots. The spectrogram for the original (unprocessed) CNC shows a clear onset burst, $f_0$ voicing, and formant curvatures. Spectrograms for vocoded stimuli show the vocoder's upper frequency limits and varying degrees of spectral and temporal resolution. The temporal energy troughs in the spectrogram for carrier H0 and the spectral energy troughs in the spectrogram for carrier S are also clearly visible. The vocoded tokens depicted in Figure 2 used Butterworth synthesis filters.

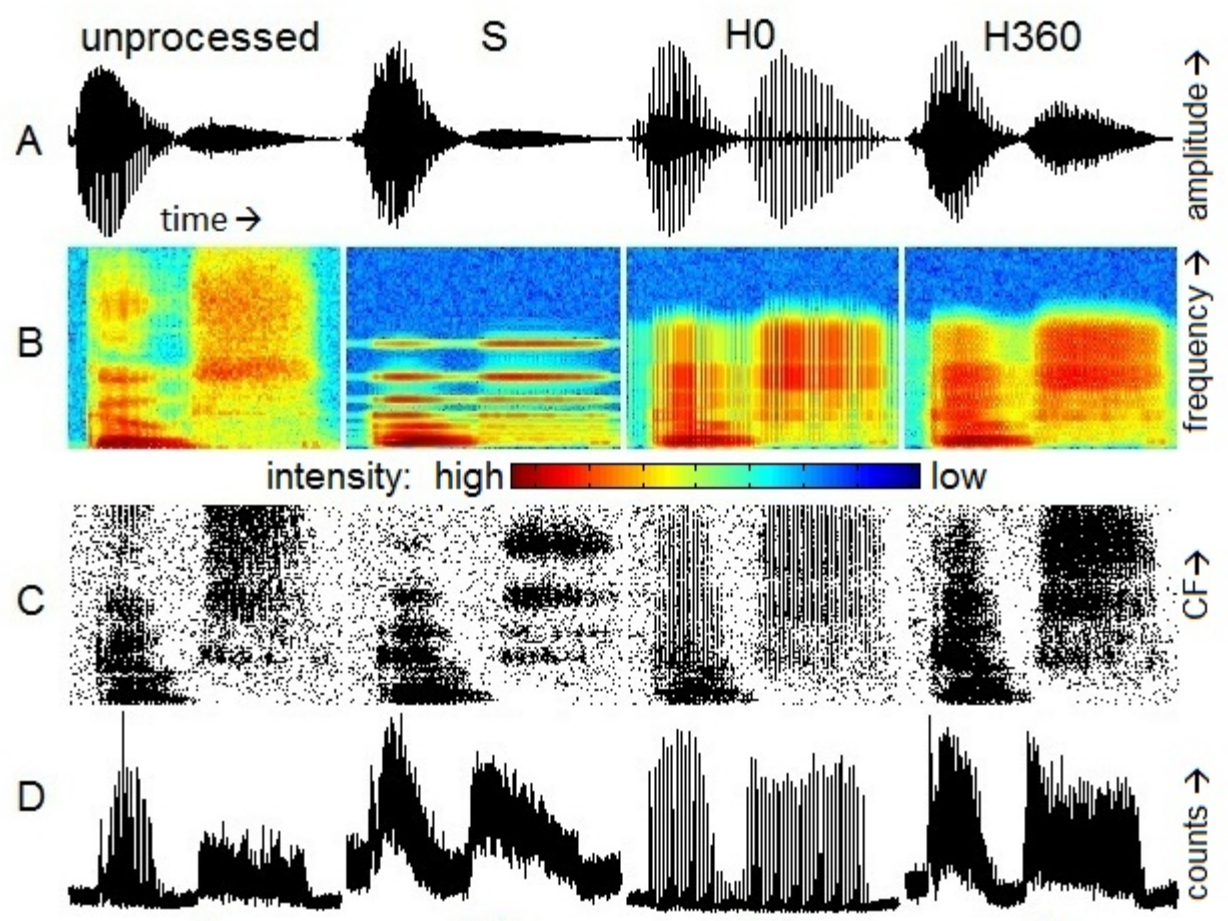



Figure 2. Plots of analyses of the CNC word "goose" in the absence of a masker allow for feature comparison among carriers, here all using Butterworth synthesis filters. Row A depicts time-domain waveforms, with intensity in arbitrary units on the vertical axis and time on the horizontal axis. Row B depicts spectrograms, with time on the horizontal axis, frequency from 0 to 10 kHz on the linear vertical axis, and intensity in arbitrary units encoded by color from blue (low) to red (high). Row C depicts modeled neural response PSTHs from 20 low-SR fibers per CF, with fiber CF ranging linearly from 0.1 to 7 kHz on the vertical axis, time on the horizontal axis, and a dot for each simulated action potential. Row D depicts summary PSTHs from 20 high-SR fibers, with counts on the vertical axis and time on the horizontal axis. The columns contain these depictions for unprocessed, S, H0, and H360 –vocoded stimuli from left to right.

**Procedure**

Twenty-three NH listeners were recruited via the University of Wisconsin's job boards, and consisted of 12 males and 11 females, aged 18-29. Listeners were screened by recording a standard audiogram, and all listeners had monaural thresholds less than 25 dB HL at all audiometric frequencies.

Thirteen listeners (7 male, 6 female) were tested on stimuli vocoded with the Butterworth synthesis filters, and 10 different listeners (5 male, 5 female) were tested on stimuli vocoded with the current spread synthesis filters. Informed consent was obtained, listeners were paid for their participation, and procedures were approved by the University of Wisconsin Human Subjects Institutional Review Board.

Closed-set, diotic, vocoded speech recognition in noise was tested in NH listeners using a forced-choice task. Stimuli were presented via headphones (Sennheiser HD600) at an average level of 60 dB A in a double-walled sound-attenuating booth (IAC). During a brief familiarization period, participants listened to each of the 50 words at least once in quiet, with each presentation vocoded with a randomly-chosen carrier. Listeners were therefore exposed to an average of fewer than 10 CNC examples of each vocoder carrier prior to testing, and were considered to be naïve rather than trained. Immediately following this exposure to vocoded speech in quiet, they listened to several of the stimuli (vocoder carrier again randomly chosen) with a background noise level of 0 dB SNR in order to acclimate to the timing of stimulus presentation within the background masker. Listeners were then tested on two blocks of 300 trials each (50 words × 6 vocoder carriers). The first block consisted of speech-in-noise stimuli with an SNR of 0 dB, and the second block consisted of stimuli with an SNR of -3 dB. Order of word and vocoder presentation was randomized for each subject, so the possible differential performance across carriers due to generalized learning (Hervais-Adelman et al. 2011) should be averaged across listeners. Testing for each block of trials lasted approximately 45 minutes and blocks were separated by a break. During each trial, the listener identified the word among the 50 CNC word choices via a computer mouse and graphical user interface, and was instructed to guess if unsure. Text representations of the words were arranged alphabetically in a 5×10 push-button matrix on the screen, and were visible throughout stimulus presentation. The user was given unlimited time to decide on his or her chosen response. No correct-answer feedback was provided during testing.

**Methods B: Phenomenological Modeling**

A computational model of the cat AN fiber (Zilany and Bruce 2006; Zilany and Bruce 2007; Zilany et al. 2009) was used to simulate responses to the vocoded and unprocessed stimuli. Shuffled cross-correlogram analyses (Joris 2003; Louage et al. 2004; Heinz and Swaminathan 2009; Swaminathan and Heinz 2012) of these simulated AN outputs were used to calculate "neural correlation coefficients," metrics of the how neural representations of envelopes and TFS of the unprocessed stimuli are preserved in the neural representations of the corresponding vocoded stimuli. These neural correlation coefficients were then entered as predictors of the psychoacoustic scores in several statistical models.

Simulated AN responses to each of the vocoded stimuli ("A"), the 50 unprocessed stimuli ("B"), and inverted versions thereof ("-A" and "-B") were generated for fibers of low, medium, and high spontaneous rates (SR), and of characteristic frequencies (CFs) of every multiple of 100 Hz from 0.1 to 7 kHz. The stimuli were resampled to 100 kHz and mathematically converted to sound pressure level values; these input values are used in the model to simulate neuronal spike post-stimulus time histograms (PSTHs). Typically when conducting correlogram analyses, sound levels are chosen independently for each fiber in order to produce the best modulation levels. However, in order to simulate actual listening conditions, a single level was chosen (60 dB) for each stimulus presentation to the model. Spikes were generated for twenty repetitions of a given stimulus, CF, and SR, and were summed to create PSTHs with 50-μs bins. The following shuffled cross-correlograms (SCCs) were calculated between pairs of stimulus PSTHs for a given fiber CF and SR: $SCC_{A/A}$, $SCC_{A/-A}$, $SCC_{B/B}$, $SCC_{B/-B}$, $SCC_{A/B}$, and $SCC_{A/-B}$. The SCC is an all-order interval histogram between all non-identical single-repetition PSTH (henceforth "$psth_i$") pairs, so refractory effects within a single model neuron are ignored. Therefore, for autocorrelograms $SCC_{A/A}$ and $SCC_{B/B}$, within-repetition intervals were subtracted from the

all-pairwise interval calculation. The convolution required for the SCC calculation was performed in Fourier space; e.g. for $SCC_{A/B}$ and $SCC_{A/A}$,

$$SCC_{A/B} = Re\left(IFT\left(FT(PSTH_A) \times FT(PSTH_B{}^*)\right)\right)$$

$$SCC_{A/A} = Re\left(IFT\left(FT(PSTH_A) \times FT(PSTH_A{}^*)\right)\right) - \sum_{i=1}^{20} Re\left(IFT\left(FT\left(psth_{A,i}\right) \times FT\left(psth_{A,i}{}^*\right)\right)\right)$$

Where * denotes complex conjugation, *Re* denotes taking the real part of the function, *FT* denotes the Fourier Transform, *IFT* denotes the Inverse Fourier Transform, $psth_i$ denotes results from the $i^{th}$ simulation of 20, and *PSTH* denotes results from the sum of the 20 simulations.

The $SCC_{A/B}$ and $SCC_{A/-B}$ cross-correlograms are representations of the similarity of the AN model response to vocoded and unprocessed stimuli. Following normalization, "sumcors" and "difcors" were calculated from pair and inverted-pair SCCs:

$$sumcor_{A/B,\ A/-B} = \frac{SCC_{A/B} + SCC_{A/-B}}{2}$$

$$difcor_{A/B,\ A/-B} = SCC_{A/B} - SCC_{A/-B}$$

The sumcor emphasizes features common to the SCC of the vocoded and unprocessed signals and the SCC of the vocoded and inverted unprocessed signals. Therefore, since envelope is thought to be independent of stimulus polarity, the sumcor is a metric that represents envelope fidelity. Likewise, the difcor emphasizes features that are different between the two SCCs. Therefore, since TFS is thought to be dependent upon stimulus polarity, the difcor is a metric of TFS fidelity. The sumcor is low-pass filtered at the fiber's CF in order to correct for "leakage" of TFS into the sumcor due to the nonlinearity of rectification present in neural responses (Heinz and Swaminathan 2009). For each stimulus and fiber

SR, the maximum values of the sumcor were averaged across fibers of all CFs, while the maximum values of the difcor were averaged across fibers of CF below 3 kHz. In order to compare across different stimuli, these averages were normalized by calculating "neural correlation coefficients" for envelope (ENV) and TFS:

$$\rho_{ENV} = \frac{sumcor_{A/B,\ A/-B}}{\sqrt{(sumcor_{A/A,\ A/-A}) \times (sumcor_{B/B,\ B/-B})}}$$

$$\rho_{TFS} = \frac{difcor_{A/B,\ A/-B}}{\sqrt{difcor_{A/A,\ A/-A} \times difcor_{B/B,\ B/-B}}}$$

Each of the vocoded stimuli therefore had a $\rho_{ENV}$ and a $\rho_{TFS}$ for each fiber SR. These neural correlation coefficients were then used as variables in linear regressions to predict psychoacoustic test scores. Additionally, $\rho$ calculations were performed within stimuli, but across CF, in order to examine temporal pattern correlation across fibers of different CF (see Swaminathan and Heinz 2011).

**Results A: Psychoacoustics**

Performance was evaluated by comparing percent correct (%C) word recognition across conditions. Figure 3 shows the average %C for each synthesis filter type and SNR as a function of vocoder carrier. Dotted lines connect the data points for harmonic complex carriers H0, H90, and H360. Error bars show 99% confidence intervals. In general, %C scores were higher with Butterworth synthesis filters and lower noise levels (0 dB SNR). For each SNR and synthesis filter combination, %C scores were highest with the H360 carrier. In contrast, worst performance was observed with carriers S and H0. Finally, performance with the harmonic complex carrier improved monotonically with increasing component phase dispersion.

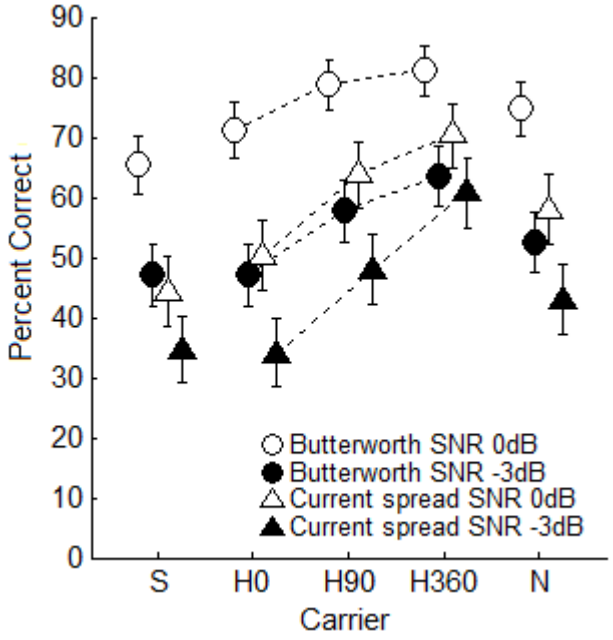



Figure 3. Psychoacoustic percent correct as a function of vocoder carrier for each of the combinations of synthesis filter and SNR, shown with 99% confidence interval error bars. Dotted lines connect the points for harmonic complex carriers with different component starting phase dispersion (H0, H90, and H360).

Individual trial response data were analyzed with a binary logistic regression in order to determine which factors were best predictors of correct responses. The full-factorial regression included carrier, synthesis filter, and SNR as categorical variables. Results revealed significant effects of carrier, synthesis filter, and SNR (all $P<0.001$), with Cox & Snell $r^2 = 0.075$. A significant interaction was also found for synthesis filter type × SNR ($P<0.001$) and synthesis filter type × carrier ($P=0.044$). An analysis of variance (ANOVA) was performed on averaged arcsine-transformed %C scores (Studebaker 1985) with carrier, synthesis filter, and SNR as factors. Results revealed significant main effects of carrier [$F_{(5,210)} = 25.0$, $P<0.001$], synthesis filter [$F_{(1,210)} = 77.0$, $P<0.001$], and SNR [$F_{(1,210)} = 124.7$, $P<0.001$], and a significant interaction of synthesis filter and SNR [$F_{(1,210)} = 7.2$, $P=0.008$]. Bonferroni-corrected post-hoc analyses showed that overall performance with the H360 carrier was significantly higher than with all carriers except H90 ($P<0.001$). Performance with the S carrier was worse than H360 ($P<0.001$), H90 ($P<0.001$), and N ($P=0.002$) carriers. Performance with the H0 carrier was worse ($P<0.001$) than with either the H360 or H90 carriers. There was no significant effect of SNR for carriers S or H360, nor was there a significant effect of synthesis filter type for carriers H90 or H360.

**Results B: Model Analyses**

Examples of PSTHs generated from the AN model for each CF and the results of summing these PSTHs across CF are shown in rows C and D of Figure 2. The model responses in row C follow the spectro-temporal patterns displayed in the corresponding waveforms and spectrograms (rows A and B). The summed PSTHs in row D reflect the broadband envelope characteristics of the time-domain signal.

Envelope and TFS neural correlation coefficients ($\rho_{ENV}$ and $\rho_{TFS}$) were averaged across words for each SNR, synthesis filter, carrier, and fiber SR, and are shown in Figure 4. As seen in the bottom row of Figure 4, averaged $\rho_{ENV}$ are positively correlated with increasing phase dispersion in the harmonic complex carriers for all fiber SRs. This reflects the trend of larger %C scores with more phase dispersion

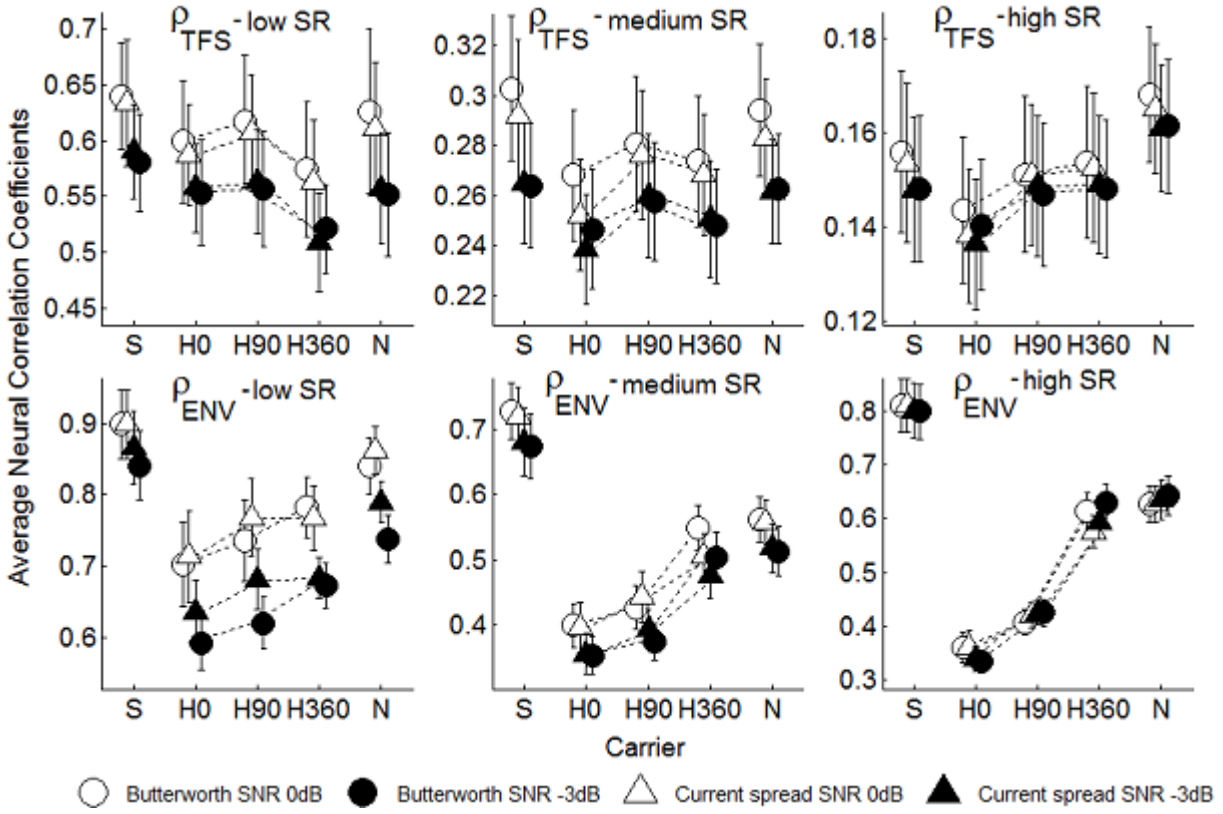



Figure 4. Neural correlation coefficients $\rho$ for TFS (top row) and envelope (bottom row) computed between vocoded stimuli and the unprocessed tokens for modeled AN fibers of low SR (first column), medium SR (second column), and high SR (third column).

as seen in Figure 3. Averaged $\rho_{ENV}$ and $\rho_{TFS}$ values are generally positively correlated with SNR, reflecting the trend of higher %C scores at the higher SNR, and indicating neural envelope and TFS information is more disrupted by higher noise levels. In contrast, the neural metrics failed to capture the performance degradation due to spectral modifications; averaged $\rho_{ENV}$ and $\rho_{TFS}$ values do not generally follow performance based on synthesis filter type. The highest averaged $\rho_{ENV}$ and $\rho_{TFS}$ values were obtained for the sine-vocoded stimuli, the stimuli with the flattest carrier envelopes. Interestingly, the sine vocoder also produced the lowest %C scores, but this may be a result of spectral rather than temporal degradations in the signal (i.e., spectral sparseness). Perhaps most striking about the neurometrics are the fairly high values of $\rho_{TFS}$. It is commonly assumed that the explicit exclusion of signals' acoustic TFS during vocoding would result in neural patterns that contain TFS information that is unrelated to that of the original acoustic waveform. However, neural TFS is defined here as that part of the neural response pattern which changes due to signal inversion. Therefore, we must conclude that the current method of vocoding leaves some of the original signal's neural TFS information intact, especially for low SR fibers. That is, the envelope representation by a vocoder preserves some aspects of the original signal that are phase-sensitive.

In order to explore the relationship between psychoacoustic performance and the effects of harmonic component phase dispersion in a simulated AN, the neural metrics were used to construct several statistical regression models. Three model classes were tested- A, B, and C. Within each model class, the predictive abilities of $\rho$ values for each SR were tested independently and together, resulting in four models per class. These models were fit to performance with the harmonic carriers only, then used to predict performance with all carriers. The models' abilities to predict performance with the harmonic complex carriers alone is also reported. Variable coefficients and statistics are shown in Table 3 (constant terms are omitted).

| Model | predictor | β | p | $R^2$ (all) | p | $R^2$ (HX) | p |
|---|---|---|---|---|---|---|---|
| A-low SR | **$\rho_{ENV}$** | **1.64** | **0.009** | 0.003 | 0.808 | 0.510 | 0.009 |
| A-medium SR | **$\rho_{ENV}$** | **1.54** | **0.009** | 0.003 | 0.820 | 0.508 | 0.009 |
| A-high SR | $\rho_{ENV}$ | 0.66 | 0.077 | 0.009 | 0.687 | 0.280 | 0.077 |
| A-all SRs | $\rho_{ENV - LOW\ SR}$ | 0.18 | 0.893 | 0.000 | 0.951 | 0.620 | 0.002 |
| | $\rho_{ENV - MED\ SR}$ | 2.90 | 0.342 | | | | |
| | $\rho_{ENV - HIGH\ SR}$ | -0.92 | 0.493 | | | | |
| B-low SR | **$\rho_{ENV}$** | **1.65** | **0.024** | 0.003 | 0.815 | 0.510 | 0.009 |
| | $\rho_{TFS}$ | -0.05 | 0.965 | | | | |
| B-medium SR | **$\rho_{ENV}$** | **0.98** | **0.012** | 0.046 | 0.363 | 0.846 | <0.001 |
| | **$\rho_{TFS}$** | **6.65** | **0.002** | | | | |
| B-high SR | $\rho_{ENV}$ | -0.05 | 0.883 | 0.124 | 0.127 | 0.619 | 0.002 |
| | **$\rho_{TFS}$** | **19.89** | **0.020** | | | | |
| B-all SRs | **$\rho_{ENV - LOW\ SR}$** | **-1.74** | **0.002** | 0.021 | 0.540 | 0.993 | <0.001 |
| | $\rho_{TFS - LOW\ SR}$ | -1.60 | 0.159 | | | | |
| | $\rho_{ENV - MED\ SR}$ | 1.04 | 0.174 | | | | |
| | **$\rho_{TFS - MED\ SR}$** | **25.41** | **<0.001** | | | | |
| | $\rho_{ENV - HIGH\ SR}$ | 1.11 | 0.063 | | | | |
| | **$\rho_{TFS - HIGH\ SR}$** | **-35.81** | **0.001** | | | | |
| C-low SR | $\rho_{ENV}$ | -8.54 | 0.727 | 0.003 | 0.825 | 0.521 | 0.008 |
| | $\rho_{TFS}$ | -12.78 | 0.676 | | | | |
| | $\rho_{TFS} \times \rho_{ENV}$ | 18.01 | 0.677 | | | | |
| C-medium SR | $\rho_{ENV}$ | 6.74 | 0.306 | 0.081 | 0.224 | 0.861 | <0.001 |
| | $\rho_{TFS}$ | 16.17 | 0.154 | | | | |
| | $\rho_{TFS} \times \rho_{ENV}$ | -22.26 | 0.376 | | | | |
| C-high SR | $\rho_{ENV}$ | -6.21 | 0.613 | 0.094 | 0.190 | 0.631 | 0.002 |
| | $\rho_{TFS}$ | 4.23 | 0.895 | | | | |
| | $\rho_{TFS} \times \rho_{ENV}$ | 41.20 | 0.616 | | | | |
| C-all SRs | $\rho_{ENV - LOW\ SR}$ | -19.23 | 0.263 | 0.032 | 0.448 | 0.997 | <0.001 |
| | $\rho_{TFS - LOW\ SR}$ | -21.84 | 0.275 | | | | |
| | $\rho_{TFS} \times \rho_{ENV - LOW\ SR}$ | 31.52 | 0.296 | | | | |
| | $\rho_{ENV - MED\ SR}$ | 29.87 | 0.274 | | | | |
| | $\rho_{TFS - MED\ SR}$ | 72.52 | 0.159 | | | | |
| | $\rho_{TFS} \times \rho_{ENV - MED\ SR}$ | -120.21 | 0.288 | | | | |
| | $\rho_{ENV - HIGH\ SR}$ | -39.72 | 0.316 | | | | |
| | $\rho_{TFS - HIGH\ SR}$ | -139.40 | 0.212 | | | | |
| | $\rho_{TFS} \times \rho_{ENV - HIGH\ SR}$ | 281.58 | 0.307 | | | | |

Table 3. Statistical model variable coefficients, correlation coefficients, and significance values for tested models, shown for prediction of performance with all carriers ("all") and with harmonic carriers only ("HX"). Significant term coefficients are shown in bold.

Model class A used the fidelity of envelope encoding ($\rho_{ENV}$) as the only predictor, first for each fiber SR separately and then for fibers of all three SRs together:

$$\%C = \beta_{i,ENV} \times \rho_{i,ENV} + \beta_{i,0}$$

$$\%C = \sum_{i=1}^{3} \beta_{i,ENV} \times \rho_{i,ENV} + \beta_0$$

where the subscript *i* indexes the fiber SRs: low (*i*=1), medium (*i*=2), and high (*i*=3). For low and medium SR fibers, positive and significant model coefficients were obtained, indicating that better envelope coding by low and medium SR fibers with dispersed-phase carriers produces improved speech recognition in noise. The all-SR model failed to produce any significant coefficients and its $\rho_{ENV}$ coefficient was negative for high SR fibers, contradicting the presumption that agreement in neural representations of vocoded and unprocessed signals should be positively correlated with psychoacoustic performance.

Although the loss of all TFS information during vocoding is generally assumed, phase-dependent response may persist. In order to assess the contributions of the fidelity of TFS encoding to speech understanding, model class B added $\rho_{TFS}$ as a predictor:

$$\%C = \beta_{i,ENV} \times \rho_{i,ENV} + \beta_{i,TFS} \times \rho_{i,TFS} + \beta_{i,0}$$

$$\%C = \sum_{i=1}^{3} \beta_{i,ENV} \times \rho_{i,ENV} + \beta_{i,TFS} \times \rho_{i,TFS} + \beta_0$$

For low and medium SR fibers, model class B produced positive and significant term coefficients for $\rho_{ENV.}$ For medium and high SR fibers, positive and significant term coefficients were obtained for $\rho_{TFS}$. These findings suggest that low SR fibers may be better at encoding what we think of as envelope

characteristics while high SR fibers may better encode TFS. The predictions due to the medium-SR model of class B and psychoacoustic results are compared in figure 5. As with the all-SR model of class A, the presence of negative coefficients in the all-SR model of class B indicates that this model is not strictly physiologically valid.

In order to further assess how the combined contributions of envelope and TFS fidelities influenced speech recognition, model class C added the interaction of $\rho_{ENV}$ and $\rho_{TFS}$ as a predictor:

$$\%C = \beta_{i,ENV} \times \rho_{i,ENV} + \beta_{i,TFS} \times \rho_{i,TFS} + \beta_{i,ENV\times TFS} \times \rho_{i,ENV} \times \rho_{i,TFS} + \beta_{i,0}$$

$$\%C = \sum_{i=1}^{3} \beta_{i,ENV} \times \rho_{i,ENV} + \beta_{i,TFS} \times \rho_{i,TFS} + \beta_{i,ENV\times TFS} \times \rho_{i,ENV} \times \rho_{i,TFS} + \beta_0$$

This is the type of model proposed in Swaminathan and Heinz (2012), although that study used only high SR fibers. No term coefficients were found to be significant with this model class.

As $\rho_{ENV}$ and $\rho_{TFS}$ values, like %C, tended to vary systematically with harmonic carrier phase dispersion, nearly all models robustly reproduced measured performance with the harmonic complex carriers.

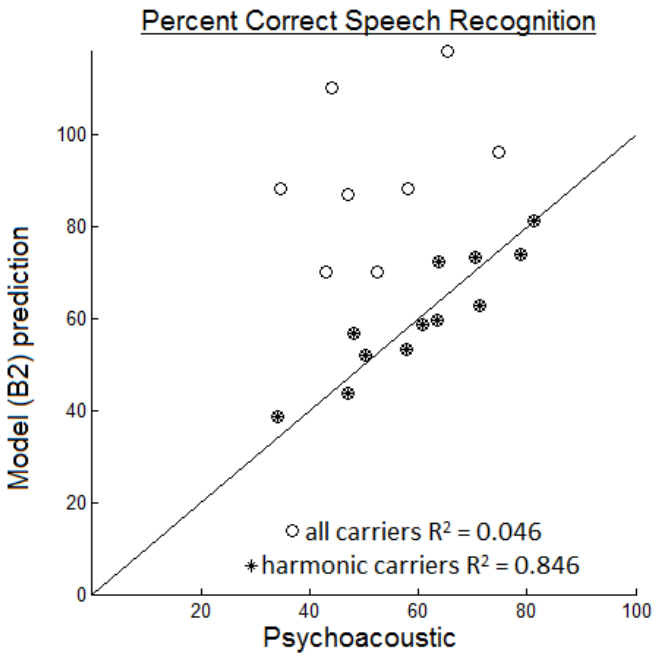



Figure 5. Measured and model-predicted percent correct speech recognition for each carrier, SNR, and synthesis filter type. The model shown here, B2, used $\rho_{TFS}$ and $\rho_{ENV}$ predictors for medium-SR fibers only and had positive and significant term coefficients.

However, as illustrated in figure 5, all models vastly over-predicted performance with the sine carrier and generally could not accurately predict performance with the noise carrier. The high $\rho_{ENV}$ values calculated for the sine vocoder indicate that it does indeed provide a flat carrier envelope for faithful signal representation, but it evidently has other characteristics that adversely affect speech understanding, such as sparse spectral representation.

We expected that the spectral profiles of the long-term AN activation patterns due to the harmonic carriers were identical due to the identity of their magnitude spectra. We also expected to observe differences among these patterns for sine, noise, and harmonic complex vocoded stimuli, and for different synthesis filter types. However, there was not a large difference in spectral profiles between stimuli with Butterworth and current spread synthesis filters in CF/counts histograms, and most vocoded stimuli tended to produce similar patterns that generally followed the corresponding unprocessed stimulus' histograms. The sine-vocoded stimulus alone showed response peaks very near the channel center frequencies at CFs > 1 kHz, i.e., the carrier frequencies, which is to be expected. This exception of the sine vocoder may be a factor contributing to its poor corresponding psychoacoustic performance. Correlation analyses of vocoded and unprocessed spectral profiles revealed no consistent patterns. Clearly, the fidelity of long-term spectra as represented in the AN could be an important factor in determining vocoded speech recognition, but that conclusion is not supported here. In order to study the spectro-temporal dynamics of evolving neuronal activity, a next step might be to analyze short time-windowed SCCs and the evolution thereof throughout a speech-like signal, but this is beyond the scope of the current analysis.

The nature of vocoders allows for speech signals to be represented by envelopes from a small number of channels. Prominent envelope fluctuations within an analysis band, perhaps generated in

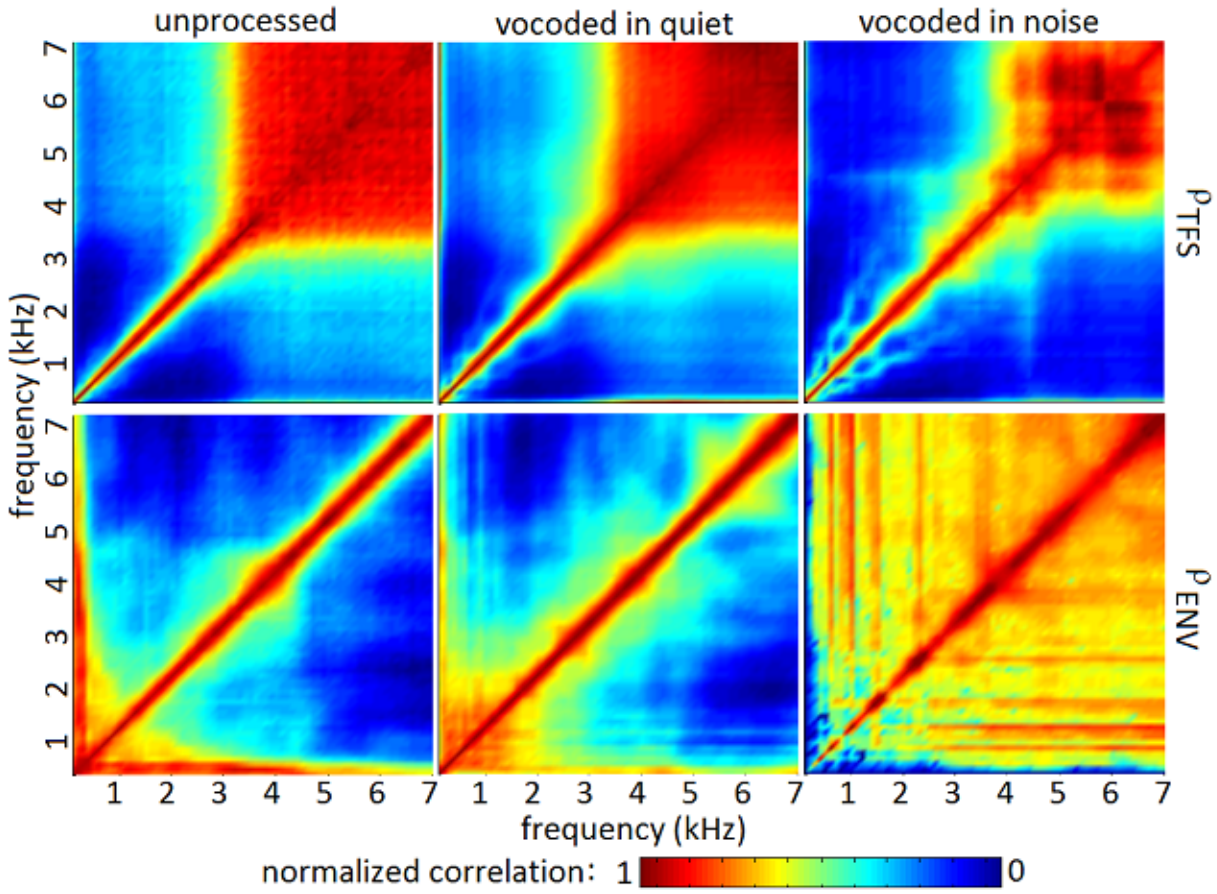



Figure 6. Across-fiber, within-stimulus neural correlation coefficients for TFS (first row) and envelope (second row). Columns show neural correlation coefficients calculated for all unprocessed stimuli, all vocoded stimuli in quiet, and all vocoded stimuli in noise from left to right. The horizontal and vertical axes depict modeled AN fiber CFs from 0.1 to 7 kHz, and color depicts the level of neural correlation from low (blue) to high (red).

very localized spectral regions, are broadcast across the entire pass-band of the vocoder's output channel. Hence, larger spectral regions of the auditory periphery are receiving coherent, smoothed, envelope information, and the loss of independent information may contribute to decreased intelligibility (Swaminathan and Heinz 2011). The across-fiber synchronous response to envelope and TFS is likely affected by vocoding and may reflect this loss of across-fiber information independence. In order to examine across-fiber envelope and TFS correlation, Figure 6 shows within-stimulus, across-CF $\rho_{ENV}$ and $\rho_{TFS}$ (as opposed to the between unprocessed and vocoded, within same CF SCCs used above)

for high SR fibers. These cross-correlations are averaged across all CNCs and conditions, and thus show the general effects of vocoding and masker addition. As expected, the peak correlation values appear along the diagonal. That is, correlations are highest along the same-CF, i.e., autocorrelation axis; in contrast, temporal firing patterns of fibers that have different CFs do not generally correlate. At CFs > 3 kHz, higher $\rho_{TFS}$ values can be seen, due to the loss of phase locking; in this case, temporal response patterns of fibers that have different CFs are no longer mathematically orthogonal. This pattern is largely consistent regardless of whether one is observing unprocessed tokens, vocoded speech in quiet, or vocoded speech in noise. These patterns are as to be expected for broadband stimuli. For unprocessed and vocoded speech in quiet, $\rho_{ENV}$ is highest among fibers of closely neighboring CF. This trend is observed for all but the lowest CFs, and means that fibers of remote CF are representing different envelopes. In contrast, for vocoded speech in noise, there is a high across-CF correlation of neural envelope representation at all CFs, indicating that redundant envelope information is carried by fibers of different CFs. The lack of unique envelope representations by different-CF fibers may be a characteristic pathology of vocoded speech in masking noise.

**Discussion**

This study examined the effect of carrier on closed-set vocoded speech recognition in noise. The carriers tested consisted of sine tones, band-pass filtered white noise, and three harmonic complexes of $f_0$ = 100 Hz with different amounts of random phase dispersion (none/0°, 90°, and 360°). Results showed that randomly dispersed starting phases resulted in improved speech intelligibility. The proposed mechanism to explain this observed trend is the improved representation of envelopes with dispersed-phase harmonic complex carriers. However, this is not well explained by neural metrics calculated from simulated AN output patterns when taking into account the also-tested sine and noise carriers. In

vocoding, the band-pass-filtered carrier is multiplied by a slowly-varying envelope calculated from the original band-pass-filtered signal for a given channel. Therefore, the output envelopes reflect the temporal characteristics of envelopes of both the input signal and the unmodulated carrier. Although they have identical magnitude spectra, the H0 and H360 carriers have very different temporal envelopes. The H0 carrier is essentially a 100-Hz biphasic pulse train, while the H360 carrier is essentially periodic noise with a repetition rate of 100 Hz. By randomizing the phases of the harmonic components in the H360 carrier, these components add to create a carrier with a flatter temporal envelope that more fully represents the signal's acoustic envelope. In low-frequency channels, the band-pass filtering renders all of the harmonic carriers very similar due to the low number of interacting harmonics within a channel. However, as the fourth channel (center frequency = 1156 Hz) is approached, a sufficient number of harmonic components are added together such that differences emerge in the carrier envelope shapes. It is the higher number of harmonic components within a single channel that causes the temporal characteristics of the signals to resemble a pulse train and periodic noise for the H0 and H360 carriers, respectively. The vocoder outputs at these higher frequency channels reflect these characteristics, which are comprised of a nearly continuous sampling of the signal envelope for carrier H360 and an envelope sampled every 10 ms for the H0 carrier. The relatively sparse envelope sampling by carrier H0 at this stage may be detrimental to speech perception even though the carrier meets the Nyquist criterion for the envelope, which was low-pass filtered at 50 Hz.

The hypothesis that temporally flatter carrier envelopes determined superior performance (e.g. Whitmal et al. 2007) posits that the acoustic information was best retained with carrier H360 due to its intrinsic temporal envelope characteristics. Examining stimuli vocoded with carriers H0 and H360 in rows B and C of Figure 2, we see that the spectral representations of the stimuli are similar, and that the vocoded stimuli differ primarily in the time domain. The flat carrier hypothesis also accounts for the

fairly high performance of the N carrier, whose envelopes are flatter for the higher-frequency, broader-band channels. However, it does not account for the high performance of carrier H90, whose temporal envelope is closer to H0 than to H360. Also, the flat carrier hypothesis does not account for the inferior performance of carrier S; with only one sine tone carrier per channel, this vocoder produced the flattest envelopes. We partially attribute the performance deficits associated with carrier S to spectral sparseness. Combined spectro-temporal effects may also be prominent factors affecting performance. It is known that stimuli with identical long-term spectra can evoke different perceptions if their temporal structures are different, and spectro-temporal patterns evoked by harmonic carriers depend largely on their phase spectra (Kohlrausch and Sander 1995; Carlyon 1996).

It is not clear from the present results how spectro-temporal pattern characteristics influence performance, but it may be instructive to inspect these patterns. Figure 7 shows the outputs of the AN model at CFs from 100 Hz to 7 kHz in 50 Hz increments for 50 ms voiced and unvoiced segments of the CNC "goose" in quiet (high SR fibers). Outputs with different carriers were compared in order to look for different spectro-temporal patterns of activation. Modeled AN fibers with CFs below ~2 kHz exhibit phase locking, which is most evident when low frequency information was present, i.e., during voiced speech. After accounting for the phase shift due to the basilar membrane traveling wave, these fibers responded largely in phase with their spectral neighbors (neighboring CFs). At higher frequencies, loss of phase locking was observed. Envelope sensitivity, as indicated by temporal bunching of fiber responses, appears to be present for CFs above ~1 kHz. It is interesting to compare model responses for voiced versus unvoiced segments among the unprocessed and vocoded tokens. The response to unprocessed speech shows clear differences between voiced and unvoiced segments; the $f_0$ is manifest by vertical striping and formant peaks are apparent in horizontal bands for the voiced segment, and a chaotically-structured high-frequency response pattern is evident for the unvoiced segment. In contrast, the

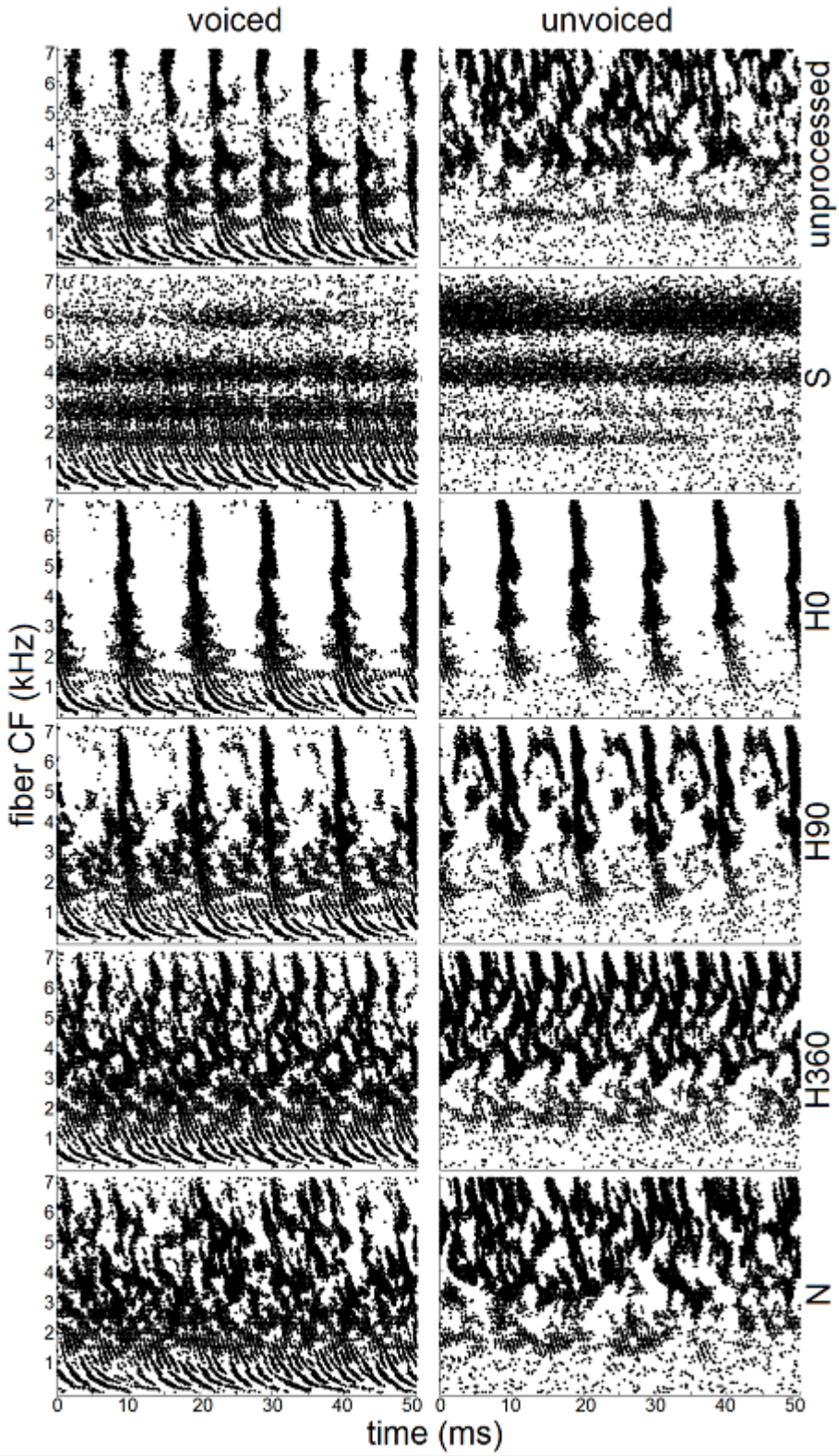



Figure 7. Model outputs for high-SR fibers with CFs between 0.1 and 7 kHz at 50 Hz spacing, shown for voiced and unvoiced 50-ms segments of the CNC “goose” in quiet, for unprocessed and vocoded stimuli.

response patterns of the vocoded speech are more similar between segments. The main difference between patterns for voiced and unvoiced segments of vocoded speech is how much energy is allocated to each band. For example, the high-frequency activation patterns for carrier H90 are similar for voiced and unvoiced segments, but the amount of energy in those high frequency bands is lower for the voiced segment. The availability of intermediate carrier envelope amplitudes to either be represented or absent in these patterns, in order to differentiate between voiced and unvoiced segments, may be a factor in the ability of a vocoder carrier to provide usable speech cues. As opposed to carriers H360 and H90, carriers S and H0 have no distinguishing temporal features which could be "turned on" as energy in a given band rises. Carrier N would have such "envelope depth" features, but they would not be consistent throughout the stimulus. However, this contrast was not investigated quantitatively. Carrier H0 resulted in a high temporal coincidence of fiber responses across CF, whereas carrier S resulted in asynchronous fiber responses across CF. The poor performance with both of these carriers indicates that temporal coincidence or asynchrony of responses of adjacent-CF fibers was not a common factor affecting performance.

The dispersed-phase harmonic carriers resulted in better speech recognition scores than carriers S or N, yet carriers S and N resulted in higher neural correlation coefficients for envelope and TFS. It is clear that many of the temporal patterns present in the acoustics were reproduced by the AN model, suggesting additional variables are needed to explain psychoacoustic performance with these vocoders. For example, the significant spectral gaps in the sine vocoder's representation of the signal could lead to a smaller number of fibers carrying that information and subsequent performance deficits. As for carrier N, since the H360 and N carriers both had relatively flat envelopes (as did carrier S), but also had the benefit of full-spectrum representation (which carrier S did not), the performance difference between

H360 and N may be due to the repetitive nature of envelope fluctuations with H360, whereas carrier N's envelope fluctuations are random.

Although the filter roll-off slopes for the Butterworth and current spread filters are roughly the same, the Butterworth filters have a flat gain in the passband, while the current spread filters are sharply peaked at the filter's center frequency. This sharpness would result in amplitude modulations becoming more quickly attenuated as the sidebands move away from the filter center frequencies. Sine-vocoded speech relies heavily on sideband detection for recognition (Souza and Rosen 2009; Kates 2011), and this loss may be responsible for the lower recognition of speech vocoded with carrier S when implementing the current spread filters. Previous literature has shown better performance with the sine vocoder when high-frequency envelope fluctuations are retained (Dorman et al. 1997; Whitmal et al. 2007; Stone et al. 2008), so the observed performance deficits with carrier S may also be due to the low (50 Hz) cutoff frequency for envelope extraction used here.

It is interesting to compare the present study's results to phenomena observed with CIs. Electric hearing largely precludes the possibility of independently-firing neighboring fibers because the current pulse phase-locks their firing (Moxon 1971; Kiang and Moxon 1972). In that respect, electric hearing seems to be much like listening with vocoder carrier H0, where fibers fire in unison, unanimously sampling and presenting the envelope at identical, discrete time points. While it has been tempting to seek to take advantage of the exquisite phase locking exhibited with electric stimulation for accurate presentation of temporal cues (van Hoesel and Tyler 2003), perhaps it is this very phenomenon which is disrupting information transfer. As illustrated by Shamma and Lorenzi (2013), internal spectrograms can be reconstituted from AN patterns by the application of a lateral inhibition network. Such inhibition mechanisms could have facilitated recovery of information not obvious in the AN patterns seen here.

Auditory nerve activation without traveling wave delays, as occurs in electric hearing, might upset such patterns of inhibition and frustrate the mechanisms that enhance internal spectrograms. However, accurate reproduction of the timing of AN activation due to traveling wave delays with electric stimulation would require extensive gains in spatial resolution relative to that with the devices commercially available today.

**References**

Bingabr, M, Espinoza-Varas, B and Loizou, P C (2008). "Simulating the effect of spread of excitation in cochlear implants." Hear Res 241(1-2): 73-79.

Carlyon, R P (1996). "Spread of excitation produced by maskers with damped and ramped envelopes." J Acoust Soc Am 99: 3647-3655.

Chen, F and Loizou, P C (2011). "Predicting the intelligibility of vocoded speech." Ear Hear 32: 331-338.

Deeks, J M and Carlyon, R P (2004). "Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation." J Acoust Soc Am 115(4): 1736-1746.

Dorman, M F, Loizou, P C and Rainey, D (1997). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs." J Acoust Soc Am 102: 2403-2411.

Dudley, H (1939). "Remaking Speech." J Acoust Soc Am 11(2): 169-177.

Eskridge, E N, Galvin, J J, Aronoff, J M, Li, T and Fu, Q J (2012). "Speech Perception with music maskers by cochlear implant users and normal hearing listeners." J Speech Lang Hear Res.

Friesen, L M, Shannon, R V, Baskent, D and Wang, X (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants." J Acoust Soc Am 110: 1150-1163.

Fu, Q J, Chinchilla, S and Galvin, J J (2004). "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users." J Assoc Res Otolaryngol 5(3): 253-260.

Fu, Q J and Nogaki, G (2004). "Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing." J Assoc Res Otolaryngol 6(1): 19-27.

Fu, Q J, Shannon, R V and Wang, X (1998). "Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing." J Acoust Soc Am 104(6): 3586-3596.

Greenwood, D D (1990). "A cochlear frequency-position function for several species - 29 years later." J Acoust Soc Am 87: 2592-2605.

Heinz, M G and Swaminathan, J (2009). "Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech." J Assoc Res Otolaryngol 10(3): 407-423.

Hervais-Adelman, A G, Davis, M H, Johnsrude, I S, Taylor, K J and Carlyon, R P (2011). "Generalization of perceptual learning of vocoded speech." J Exp Psychol Hum Percept Perform 37(1): 283-295.

Joris, P X (2003). "Interaural time sensitivity dominated by cochlea-induced envelope patterns." J Neurosci 23(15): 6345-6350.

Kates, J M (2011). "Spectro-temporal envelope changes caused by temporal fine structure modification." J Acoust Soc Am 129: 3981-3990.

Kiang, N Y and Moxon, E C (1972). "Physiological considerations in artificial stimulation of the inner ear." Ann Otol 81(5): 714-730.

Kohlrausch, A and Sander, A (1995). "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets." J Acoust Soc Am 97(3): 1817 - 1829.

Loizou, P C (2006). Speech processing in vocder-centric cochlear implants. Cochlear and Brainstem Implants. A. Moller. Basel, Karger. 64: 109-143.

Louage, D H, van der Heijden, M and Joris, P X (2004). "Temporal properties of responses to broadband noise in the auditory nerve." J Neurophysiol 91(5): 2051-2065.

Lu, T, Carroll, J and Zeng, F G (2007). On acoustic simulations of cochlear implants. Conference on Implantable Auditory Prostheses, Lake Tahoe, CA.

Moxon, E C (1971). Neural and mechanical responses to electrical stimulation of the cat's inner ear. Dissertation, Massachusetts Institute of Technology.

Nelson, P B, Jin, S-H, Carney, A E and Nelson, D A (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners." J Acoust Soc Am 113: 961-968.

Schroeder, M R (1966). "Vocoders: Analysis and Synthesis." Proc IEEE 54: 720-734.

Schroeder, M R (1970). "Synthesis of low peak-factor signals and binary sequences with low autocorrelation." IEEE Trans Inf Theory: 85 - 89.

Shamma, S and Lorenzi, C (2013). "On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system." J Acoust Soc Am 133(5): 2818-2833.

Shannon, R V, Zeng, F-G, Kamath, V, Wygonski, J and Ekelid, M (1995). "Speech recognition with primarily temporal cues." Science 270(5234): 303-304.

Souza, P and Rosen, S (2009). "Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech." J Acoust Soc Am 126(2): 792-805.

Stone, M A, Füllgrabe, C and Moore, B C J (2008). "Benefit of high-rate envelope cues in vocoder processing: effect of number of channels and spectral region." J Acoust Soc Am 124: 2272-2282.

Studebaker, G A (1985). "A "rationalized" arcsine transform." J Speech Hear Res 28: 455 - 462.

Swaminathan, J and Heinz, M G (2011). "Predicted effects of sensorineural hearing loss on across-fiber envelope coding in the auditory nerve." J Acoust Soc Am 129(6): 4001-4013.

Swaminathan, J and Heinz, M G (2012). "Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise." J Neurosci 32(5): 1747-1756.

Van Deun, L, Van Wieringen, A and Wouters, J (2010). "Spatial Hearing Perception Benefits in Young Childgren With Normal Hearing and Cochlear Implants." Ear Hear 31(5): 702-713.

van Hoesel, R J M and Tyler, R S (2003). "Speech perception, localization, and lateralization with bilateral cochlear implants." J Acoust Soc Am 113(3): 1617-1630.

Whitmal, N A, Poissant, S F, Freyman, R L and Helfer, K S (2007). "Speech intelligibility in cochlear implant simulations : Effects of carrier type , interfering noise , and subject experience." J Acoust Soc Am 122(4): 2376 - 2388.

Zilany, M S and Bruce, I C (2006). "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery." J Acoust Soc Am 120(3): 1446.

Zilany, M S and Bruce, I C (2007). "Representation of the vowel /epsilon/ in normal and impaired auditory nerve fibers: model predictions of responses in cats." J Acoust Soc Am 122(1): 402-417.

Zilany, M S, Bruce, I C, Nelson, P C and Carney, L H (2009). "A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics." J Acoust Soc Am 126(5): 2390-2412.

# Chapter 3

# Exploring location cue encoding in the auditory periphery with a computational model

**Introduction**

Much research has been performed attempting to understand how the neural signals produced in the peripheral auditory system are processed by binaural neural machinery to produce sounds' location-dependent percepts. The physical cues regarding a sound's azimuthal location are due to listeners having two sound-receptive organs physically separated by a head. The physical separation of the ears facilitates the introduction of interaural phase or time differences (ITDs) between the ears for non-zero source azimuths. The head itself produces an acoustic shadow which results in azimuth- and wavelength-dependent interaural level differences (ILDs). The venerable Duplex Theory of sound localization holds that ITDs are the dominant lateral location cue for low frequencies, while ILDs are

more dominant for high frequencies (Strutt, 2007; Macpherson & Middlebrooks, 2002). Additionally, two different neural pathways are thought to be principally responsible for the binaural processing of these cues in mammals; ILDs and envelope ITDs are generally considered to be processed in the lateral superior olive (LSO), and fine structure ITDs are considered to be primarily processed in the medial superior olive (MSO) (Goldberg & Brown, 1969; Grothe, Pecka, & Mcalpine, 2010; Yin & Chan, 1990).

The general mechanisms whereby sound is transduced in the auditory periphery, resulting in firing patterns of the auditory nerve (AN), are perhaps better understood than their links to actual percepts. Sound intensity is indicated by higher levels of activity on the AN. Frequency is encoded in the place of maximum displacement of the basilar membrane (BM) and/or by the temporal patterns of AN firings which phase-lock to its low frequency oscillations. The frequency to which a given AN fiber responds best is called its characteristic frequency (CF). In the binaural processing pathway, ILDs are thought to be encoded from differential activity levels in nerve fibers arriving from the cochlear nucleus (CN) by contralateral inhibition and ipsilateral excitation in the LSO. The mechanism of coding ITD cues in the midbrain is still unclear, but is known to rely in part upon the differential timing of action potentials in fibers from the left and right CN. This required fine-timing information is initiated by the phase-locking of AN firing patterns to sounds' temporal fine structure (TFS) as represented by BM motion at frequencies below several kHz. These phase-locked signals in the AN are further temporally sharpened by bushy cells in the CN. Additional ITD cues carried in sounds' slowly-varying envelopes may be represented by differential neuronal activity levels and/or individual spike timing.

The classical model of ITD processing is by Jeffress (1948), and relies on delay lines to facilitate ITD-dependent coincidence detection. This is usually represented simply by the calculation of cross-correlation of signals from the two ears. In the Jeffress model, fibers of the same CF from each ear's CN

synapse onto multiple coincidence-detecting cells, each with a different inherent relative delay between signals coming from the left and right ears. When a given ITD exactly compensates for one of these cells' intrinsic left/right delays, coincidence is detected in that cell and it fires, indicating that given ITD was detected. The prime attractiveness of the Jeffress model lies in its simplicity and ease of computation. However, although the neural machinery of coincidence detection and sensitivity to "best delays" is known to exist (Rose et al., 1966), physiologic evidence of axonal delay lines in mammals is lacking. Another binaural mechanism, first proposed by Schroeder (1977), and further developed by Shamma (1989), uses the delays inherent in the cochlea for its cross-correlation computation. This "stereausis" model computes the cross-correlation from signals in fibers of different CFs between left and right ears, the difference in CF being proportional to an inherent relative delay due to cochlear mechanics. Bonham & Lewis (1999) found this cochlear delay source to be sufficient for ITD sensitivity. A third possible source of binaural delay relies on inhibition (Brand, Behrend, Marquardt, McAlpine, & Grothe, 2002). Brughera, Dunai, & Hartmann (2013) tested the ability of several models of the MSO to reproduce psychophysical sine tone discrimination thresholds, and finding no "best" model, proposed that a combination of mechanisms is responsible. Generally, these models are designed primarily to deal with on-going and static ITDs. Onset and offset ITDs also provide prominent perceptual cues to location, and dynamically-changing ITDs result in weak, or "sluggish" perceptual cues (Stern & Trahiotis, 1995). These models may therefore have difficulty predicting the outcomes for complex signals such as speech.

Much research has investigated the psychophysical and neurophysiological differences due to sounds' envelopes and TFS. When applied to an acoustic signal, these two terms refer to the magnitude and phase outputs of a modulation filtering decomposition such as the Hilbert transform. In the context of neural signals, the terms envelope and TFS usually refer to those temporal aspects of the action potential firing patterns which are unaffected by or are largely affected by an inversion of the input

acoustic signal, respectively. Neural TFS, therefore, is defined as that pattern of neural response which is affected by phase locking, and is therefore critical for those mechanisms in the auditory system which are thought to rely on precisely-timed action potentials, such as ITD sensitivity. Therefore, this envelope/TFS dichotomy is also very relevant in investigations of binaural hearing mechanisms.

The coding of neural TFS and envelope has been investigated by shuffled correlogram analyses of recorded (Joris, 2003; Louage, van der Heijden, & Joris, 2004) and simulated fiber response data (Heinz & Swaminathan, 2009; Swaminathan & Heinz, 2011, 2012). The present study extended these analyses and others to explore theoretical binaural mechanisms for encoding location cues. The same model as used by Heinz and Swaminathan was used to generate simulated AN response data (Zilany, Bruce, Nelson, & Carney, 2009; Zilany & Bruce, 2006, 2007). From simulated AN data, several metrics were calculated for the prediction of azimuthal location and then assessed for their ability to predict psychoacoustic results of a virtual acoustic space (VAS) localization study using vocoded and unprocessed (or "clean") speech tokens.

**Methods**

### A. Nonspeech stimuli

Twelve different one-second acoustic stimuli, consisting of clicks, tones, noise, and harmonic complexes, were processed by the Zilany AN model. Stimuli and conditions are shown in Table 1. Stimuli contained azimuth cues for horizontal plane locations from -90° to +90° in 5° increments. Azimuth cues were generated from the KEMAR head-related transfer functions (HRTFs) and consisted of either 1) all HRTF cues, 2) ILDs and spectral shift only (zero ITD), or 3) ITDs only. Medium and high spontaneous rate (SR) fibers of 100 different CFs logarithmically spaced between 100 Hz and 16 kHz were simulated. Stimuli were on- and off-ramped using halves of a 100-ms Blackman window and were presented to the

| Acoustic Stimuli | Spontaneous Rates | Locations | Metrics |
|---|---|---|---|
| • + click<br>• +/- click<br>• -/+ click<br>• White noise<br>• 100 Hz sine tone<br>• 200 Hz sine tone<br>• 500 Hz sine tone<br>• 1.5 kHz sine tone<br>• 3 kHz sine tone<br>• 5 kHz sine tone<br>• 100 Hz sine phase harmonic complex<br>• 100 Hz random phase harmonic complex | • Medium<br>• High<br><br>Azimuth Cues<br>• Full HRTF<br>• ILD and spectral tilt<br>• ITD<br><br>Polarities<br>• Normal<br>• Inverted | • -90°<br>• -85°<br>• -80°<br>.<br>.<br>.<br>• 0°<br>.<br>.<br>.<br>• 80°<br>• 85°<br>• 90° | • ILD<br>• Time delay of peak in same-CF for:<br>• cross-correlation<br>• difcor<br>• sumcor<br>• Off-center diagonal of peak in sum of binaural correlation plane for:<br>• inner product<br>• difcor<br>• sumcor |

Table 1. Auditory nerve responses were simulated for left and right ear channels of twelve acoustic stimuli at 37 azimuthal locations in the frontal horizontal plane. Fibers of 2 SRs and 100 CFs logarithmically spaced between 100 Hz and 16kHz were used. Location cues consisted of full HRTF cues, ITDs only, or ILDs only. Seven metrics were extracted in order to examine possible neural mechanisms for lateralization.

model at a level of 60 dB SPL. For original and polarity-inverted versions of the left- and right-ear signals for each stimulus, twenty repetitions of the AN model simulation were performed. Post-stimulus time histograms (PSTHs) were constructed for each fiber by arranging simulated spike times into 50-μs bins. These left- and right-ear PSTHs were used in the calculation of various metrics to predict the azimuths from which they were generated.

For extracting ITDs from simulated AN patterns, the most straightforward calculation available, and the one giving the best resolution here (50 μs), is the cross-correlation of PSTHs from left and right ears of the same CF. The time delay (abscissa) of the peak in this cross-correlation function strongly reflects the difference in sound arrival times at each ear for frequencies below which ILDs are negligible. (Different levels at the ears could result in different phases at which a waveform is transduced.) This metric is therefore used as a standard benchmark for control. The all-CF sum of cross correlations between PSTHs from left- and right-ear simulated fibers of the same CF are calculated thusly:

$$XC(L,R) = \sum_{i=1}^{100} Re\left(IFT\left(FT\left(PSTH_{i,L}\right) \times FT\left(PSTH_{i,R}{}^{*}\right)\right)\right)$$

where * denotes complex conjugation, *Re* denotes taking the real part of the function, *FT* denotes the Fourier Transform, *IFT* denotes the Inverse Fourier Transform, and $PSTH_{i,L\ or\ R}$ denotes the sum of 20 post-stimulus time histogram responses of a fiber of CF *i* to stimulus presentations in the left or right ear. The time delay abscissa associated with the peak of this function is the extracted metric:

$$x.XC = abscissa\ of\ peak\ in\ XC(L,R)$$

Similarly, peak abscissae are extracted as metrics from the "difcor" and "sumcor" combinations of cross-correlation functions. These two functions represent common TFS- and ENV-sensitive responses between simulated firing patterns in the two fibers compared. The *difcor* is defined as the difference between the cross-correlation of two signals R and L and the cross-correlation of R with polarity-inverted L, $\bar{L}$:

$$x.dif = abscissa\ of\ peak\ in\ \left(XC(L,R) - XC(\bar{L},R)\right),$$

and the *sumcor* is defined as the average of the two cross correlations:

$$x.sum = abscissa\ of\ peak\ in\ \frac{1}{2}\left(XC(L,R) + XC(\bar{L},R)\right)$$

The physiological execution of such calculations as determining the "best delay" requires delaying mechanisms or delay lines. However, conclusive anatomical evidence for an elaborate network of delay lines lacking, we focus on ITDs extracted from a cross-CF coincidence-detection network, i.e., the stereausis network. The delays here are provided by the cochlea, and the temporal resolution of these delays is determined by the density of CFs of fibers analyzed. Although it is known that signals on

the AN are enhanced by cells in the CN prior to the information reaching the superior olivary complex, the involved mechanisms are not entirely clear. Thus, we have omitted such enhancements as the lateral inhibition network of Shamma (1985, 1989). In summary, the calculations undertaken here to predict lateral positions are based solely on AN patterns and the assumption of simple binaural coincidence detection machinery.

Three zero-internal-delay cross-correlation measures were obtained for all fiber CF pairs. First, the *inner product* of all PSTH pairs was calculated, giving a direct measure of cross correlation between AN patterns in all fiber pairs with only cochlear delays. For this calculation, cross-correlations were performed between all left/right PSTH pairs. The inner product for each CF pair is the value of this cross-correlation at zero delay. For *sumcor* and *difcor* measures, cross correlations were also calculated between unmodified and one-ear-polarity-inverted stimuli, as above, and the values at zero delay for each CF pair calculation were recorded. These three zero-delay binaural cross-correlation measures- inner product, difcor, and sumcor- are described in detail below. Individual PSTHs were discarded immediately following the extraction of desired measures due to the large amount of data involved; 12 signals × 37 locations × 3 cue sets × 2 fiber SRs × 100 left CFs × 100 right CFs × 2 polarities × 20 repetitions > 1 billion simulations. Indeed, the computational enterprise depended entirely upon the use of the many cores available on the Open Science Grid.

Symbols $\mathbf{L}_i$ and $\mathbf{R}_j$ represent the PSTHs for the $i^{th}$ CF fiber in the left-ear channel and the $j^{th}$ fiber in the right-ear channel, respectively. The symbol $\bar{\mathbf{L}}_i$ represents the PSTH generated from the inverted left-ear waveform. The *inner product* $\mathbf{<L_i|R_j>}$, the cross-correlation of $\mathbf{L}_i$ and $\mathbf{R}_j$ at zero time delay, is the sum of the product of the bin values in the left and right PSTHs:

$$\langle L_i | R_j \rangle = \sum_{k=all\ bins} PSTH_{i,L}(k) \times PSTH_{j,R}(k)$$

The *difcor*, $\mathbf{\langle L_i | R_j \rangle - \langle \bar{L}_i | R_j \rangle}$**,** measures the difference between the correlation of left and right PSTHs due to the original signals and the correlation of left and right PSTHs when the left-ear channel PSTH is generated by an inverted waveform. These are calculated as the difference between cross-correlation functions at zero time delay. If this difference is large, then the two left-ear AN patterns (due to original and inverted waveforms) were significantly dependent upon the waveform phase, i.e., TFS. The sign of this difference indicates the relative phase difference between the waveforms at the two different fibers. For instance, a signal with zero ITD and sufficient low-frequency content for phase locking to TFS will produce roughly identical signals in **L** and **R** at a given CF while signals in $\bar{\mathbf{L}}$ and **R** may be antiparallel at low CFs, where phase locking is strong. At a certain nonzero ITD, however, the difference in response timing between **L** and **R** will cause *those* signals to be out of phase, while $\bar{\mathbf{L}}$ and **R** correlate highly. Thus, this definition of difcor represents a correlation of binaural response to TFS. If the difference is low, there is little TFS sensitivity.

The *sumcor*, $\mathbf{[\langle L_i | R_j \rangle + \langle \bar{L}_i | R_j \rangle]/2}$**,** measures the mean of the correlation of left and right PSTHs due to original signals and the correlation when the left-ear signal is inverted. Whereas the difcor represents a measure of binaural correlation that is sensitive to TFS, this sumcor represents a measure of binaural correlation that is *insensitive* to TFS. To say it is purely sensitive to envelope is not strictly correct, as the mean response due to all phase shifts would provide a truer measure:

$$sumcor \sim \sum_{\theta=0}^{2\pi} \langle L_{i,\theta} | R_j \rangle$$

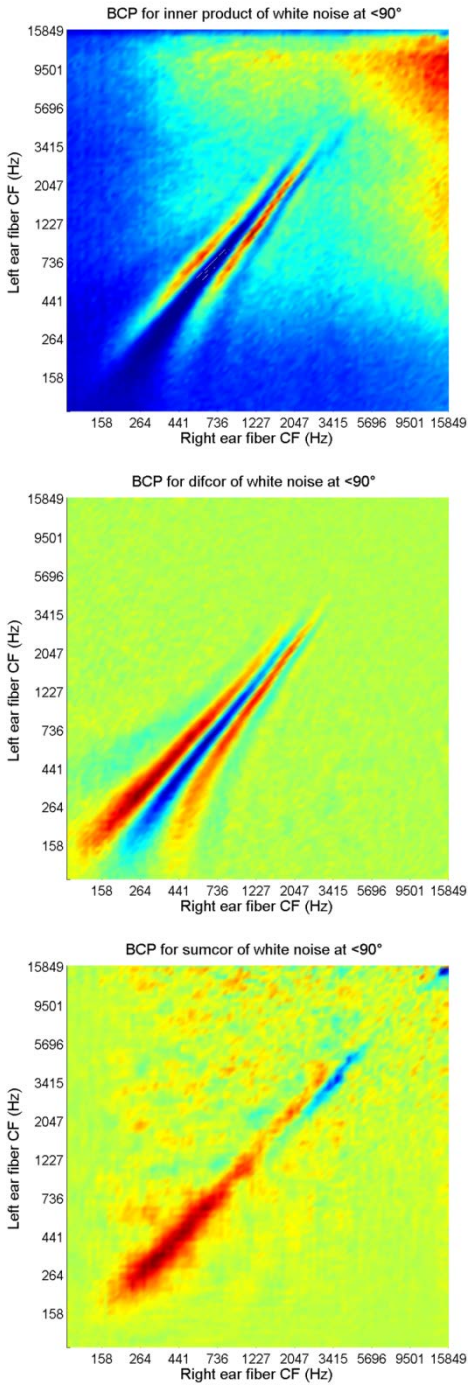



Figure 1. Binaural correlation plane due to the inner product, difcor, and sumcor calculations for the white noise nonword stimulus at azimuth +90°. The heat map shows high to low values from red to green to blue. Each panel has a unique color axis.

However, the artificiality of including only two phases in this average is partially whitewashed by the following procedure. The sumcor cross-correlation functions are "corrected" by low-pass filtering it at the fiber CF, which removes TFS "leaked" into higher frequencies due to the rectification nonlinearity in transduction. In practice, this low-pass filtering is executed in the frequency domain for each ear (since they may be fibers of different CF for each ear) during the cross-correlation calculations. This estimation of envelopes is equivalent to full-wave rectification and low-pass filtering.

The above three zero-internal-delay cross-correlation measures- inner product, difcor, and sumcor- are calculated for each of 100×100 CF pairs for a given stimulus, resulting in a "binaural correlation plane," or BCP. The extraction of predictive metrics from these BCPs is described further below. Example BCPs are plotted as "heat maps" in Figure 1. These plots depict values of each measure as a color for each CF pair, red indicating high values and blue indicating low values. The color axis runs from blue through green to red and spans the range of values on a given BCP from the minimum to the maximum. Therefore, no two BCPs' color axes are necessarily the same. For example, on difcor BCPs, blue normally indicates

negative values, red normally indicates positive values, and green normally indicates values near zero. In contrast, on a sumcor BCP, nearly all values are close to 1, and colors may represent a range between 0.9 and 1.

The stereausis model posits that delays required for cross-correlation analyses with coincidence detecting neurons are those frequency-dependent delays associated with cochlear transduction due to the mechanical properties of the basilar membrane (BM). The cochlea's tonotopicity is monotonic and results in a monotonic dependence of time delay upon frequency. A peak in the correlation of signals in fibers of different CFs therefore indicates time coincidence of transduction. The difference between these fiber CFs subsequently indicates the ITD, given the difference in traveling wave travel time between the two fibers' physical locations on the BM. The detection of coincidence between fibers of different CFs could easily be implemented with simple neural machinery. For a signal with a given ITD, these coincidences would occur at different-CF fiber pairs along the length of the signal's bandwidth, and could be measured with an array of coincidence detection cells innervated by these specific pairs. Therefore, a given ITD would be represented as a unique *diagonal line* along the BCP, an ITD of zero being represented by a line through the $CF_{left} = CF_{right}$ diagonal. In order to determine along which diagonal line this peak in the BCP lies, the BCPs are summed along the direction of the $CF_{left} = CF_{right}$ diagonal axis and normalized by the number of fiber comparisons (e.g., along the $CF_{left} = CF_{right}$ axis, there are 100 comparisons, so the sum is divided by 100, whereas the $CF_{left,second\ lowest}$ x $CF_{right,highest}$ + $CF_{left,lowest}$ x $CF_{right,second\ highest}$ diagonal comprises only two comparisons and thus divided by 2). The inner product BCP sum, *IP,* is thus

$$IP(j = 1\ to\ 100) = \frac{1}{j}\sum_{i=1}^{j}\langle L_i | R_{100-j+i}\rangle$$

$$IP(j = 100\ to\ 199) = \frac{1}{200 - j} \sum_{i=1}^{200-j} \langle R_i | L_{j-100+i} \rangle$$

Likewise for difcor and sumcor BCP summation measures, *dif* and *sum*:

$$dif(j = 1\ to\ 100) = \frac{1}{j} \sum_{i=1}^{j} \langle L_i | R_{100-j+i} \rangle - \langle \bar{L}_i | R_{100-j+i} \rangle$$

$$dif(j = 100\ to\ 199) = \frac{1}{200 - j} \sum_{i=1}^{200-j} \langle R_i | L_{j-100+i} \rangle - \langle R_i | \bar{L}_{j-100+i} \rangle$$

$$sum(j = 1\ to\ 100) = \frac{1}{2j} \sum_{i=1}^{j} \langle L_i | R_{100-j+i} \rangle + \langle \bar{L}_i | R_{100-j+i} \rangle$$

$$sum(j = 100\ to\ 199) = \frac{1}{400 - 2j} \sum_{i=1}^{200-j} \langle R_i | L_{j-100+i} \rangle + \langle R_i | \bar{L}_{j-100+i} \rangle$$

The extracted metrics are the off-center diagonal number abscissae associated with the peak of these functions:

$$s.IP = abscissa\ of\ peak\ in\ IP$$

$$s.dif = abscissa\ of\ peak\ in\ dif$$

$$s.sum = abscissa\ of\ peak\ in\ sum$$

The final extracted metric is the ILD. Shamma (1989) used the stereausis correlation function $c_{ij} = (x_i^2 + y_j^2)$, where x and y are signals from the left and right fibers, whereas we use a pure coincidence detection of the form $c_{ij} = x_i y_j$. As noted by Shamma (1989), this coincidence function does not explicitly

produce lateral measures of ILD. Here, calculation of the ILD metric consists simply of finding the difference between the numbers of spikes simulated for left and right fibers:

$$ILD = \sum_{i=1}^{100} PSTH_{i,R} - \sum_{i=1}^{100} PSTH_{i,L}$$

In summary, the seven metrics consisted of neural ILDs, the time delays of peaks in the sums across all 100 CFs of same-CF cross-correlation functions, difcors, and sumcors, and the locations of peaks in the diagonal sums of BCPs using inner product, difcor, and sumcor calculations. These seven metrics are shown in the rightmost column of Table 1, and were used as factors, *p*, in models used to predict generating lateral azimuths, $\theta$, by the linear function

$$\theta = \beta \times p(\theta) + constant$$

Such a function was built for each of the stimuli and set of predictors using extracted metric *p* at each azimuth $\theta$ from -90° to 90° in 5° increments. This was done for each location cue set and for medium and high SR fibers.

It may be that instead of rows of coincidence detection neurons tuned to specific ITDs, the physiological calculation of ITDs relies on a procedure more akin to template pattern matching. Therefore, an additional set of models were developed wherein the entire BCPs themselves were used as predictors. These calculations used only high SR fibers and stimuli with full HRTF cues, and were performed for azimuths from -90° to 90° in 10° increments. For the inner product, an angle φ between pairs (a,b) of stimulus and azimuth combinations was defined by:

$$\text{IP: } \cos\varphi_{(a,b)} = 2\sum_{i=1}^{100}\sum_{j=1}^{100}\frac{\langle L_i|R_j\rangle_a \times \langle L_i|R_j\rangle_b}{\langle L_i|R_j\rangle_a^2 + \langle L_i|R_j\rangle_b^2} = 2\frac{\langle \boldsymbol{L}|\boldsymbol{R}\rangle_a \times \langle \boldsymbol{L}|\boldsymbol{R}\rangle_b}{\langle \boldsymbol{L}|\boldsymbol{R}\rangle_a^2 + \langle \boldsymbol{L}|\boldsymbol{R}\rangle_b^2}$$

where $\langle \boldsymbol{L}|\boldsymbol{R}\rangle_x$ denotes the entire inner product BCP. Similar definitions follow for sumcor and difcor representations:

$$\text{difcor: } \cos\varphi_{(a,b)} = 2\frac{(\langle \boldsymbol{L}|\boldsymbol{R}\rangle_a - \langle \bar{\boldsymbol{L}}|\boldsymbol{R}\rangle_a) \times (\langle \boldsymbol{L}|\boldsymbol{R}\rangle_b - \langle \bar{\boldsymbol{L}}|\boldsymbol{R}\rangle_b)}{(\langle \boldsymbol{L}|\boldsymbol{R}\rangle_a - \langle \bar{\boldsymbol{L}}|\boldsymbol{R}\rangle_a)^2 + (\langle \boldsymbol{L}|\boldsymbol{R}\rangle_b - \langle \bar{\boldsymbol{L}}|\boldsymbol{R}\rangle_b)^2}$$

$$\text{sumcor: } \cos\varphi_{(a,b)} = 2\frac{(\langle \boldsymbol{L}|\boldsymbol{R}\rangle_a + \langle \bar{\boldsymbol{L}}|\boldsymbol{R}\rangle_a) \times (\langle \boldsymbol{L}|\boldsymbol{R}\rangle_b + \langle \bar{\boldsymbol{L}}|\boldsymbol{R}\rangle_b)}{(\langle \boldsymbol{L}|\boldsymbol{R}\rangle_a + \langle \bar{\boldsymbol{L}}|\boldsymbol{R}\rangle_a)^2 + (\langle \boldsymbol{L}|\boldsymbol{R}\rangle_b + \langle \bar{\boldsymbol{L}}|\boldsymbol{R}\rangle_b)^2}$$

where $\langle \boldsymbol{L}|\boldsymbol{R}\rangle_x \pm \langle \bar{\boldsymbol{L}}|\boldsymbol{R}\rangle_x$ denotes the sumcor or difcor BCP. Thus the angle φ between correlation plane patterns due to any pair of inner product/difcor/sumcor – stimulus – azimuth combinations (a,b) was calculated. By observing how these angles changed as a function of azimuths compared for a given stimulus and measure, the "uniqueness" of a BCP pattern for a given azimuth could be ascertained. When the pair of BCPs compared were identical, i.e., same azimuth and stimulus, the resulting angle is zero. When the BCP patterns were different, an angle between 0° and 90° resulted. It may be instructive to quantify the "uniqueness" of these BCP pattern cues for a given azimuth by comparing the cos(φ) values along the same-azimuth diagonal with those cos(φ) values off the same-azimuth diagonal. A uniqueness angle, δ, is thereby defined by:

$$\cos(\delta) = \frac{1}{342}\sum_{\theta_1=-90^\circ}^{90^\circ}\sum_{\theta_2\neq\theta_1} cos(\varphi_{\theta_1\theta_2}) - \frac{1}{19}\sum_{\theta=-90^\circ}^{90^\circ}\cos(\varphi_{\theta,\theta})$$

However, the precisely unity values of cos(φ) along the same-azimuth diagonal is an artificiality of precisely identical BCPs, these BCPs being generated from identical, stochastically-generated PSTHs. Therefore, an additional measure was made wherein diagonal cos(φ) values were interpolated from immediately adjacent off-diagonal elements. This resulted in a "blunting" of the diagonal measure, giving the blunted uniqueness angle, $\delta_b$, by:

$$\cos(\delta_b) = \frac{1}{342} \sum_{\theta_1=-90^\circ}^{90^\circ} \sum_{\theta_2 \neq \theta_1} cos(\varphi_{\theta_1\theta_2}) - \frac{1}{36} \sum_{\theta=-80^\circ}^{90^\circ} \cos(\varphi_{\theta,\theta-10^\circ}) - \frac{1}{36} \sum_{\theta=-90^\circ}^{80^\circ} \cos(\varphi_{\theta,\theta+10^\circ})$$

This blunted uniqueness angle was computed for inner product, difcor, and sumcor measures for each stimulus.

## B. CNC Stimuli

The same simulations and calculations as above were conducted on unprocessed and noise-vocoded VAS stimuli consisting of 5 consonant-nucleus-consonant words (CNCs; “beam”, ”cape”, “car”, “choose”, “chore”) at 19 different azimuths (-90° to +90° in 10° increments). These calculations used high SR fibers only. The VAS stimuli were created from the recorded HRTF of a listener who completed a localization study with these stimuli. Models were used to predict generating azimuth, as above, but also the listener’s psychophysical responses, providing insight regarding the relative importance of TFS and envelope cues as represented in AN patterns for lateralization. Models’ metric coefficients were generated by least-squares fits of metrics from clean VAS stimuli to the relevant azimuth. Correlation coefficients were calculated for predictions of both clean and vocoded VAS stimulus azimuths. Another model was generated that used metric coefficients derived from the 12 nonword stimuli above and was tested for both clean and vocoded VAS stimuli.

As with the nonword stimuli, BCP patterns were used to predict lateralization of clean and vocoded VAS stimuli, and an angle φ was calculated for each azimuth and stimulus pair for inner product, sumcor, and difcor BCPs. These azimuth plane patterns in cos(φ) were also analyzed using the blunted uniqueness angle measure $\delta_b$. Considering the nature of BCP patterns due to the vocoded stimuli, the importance of blunting the azimuth-dependent φ patterns is paramount. Without doing so,

pairs of vocoded stimuli at identical azimuths artificially look singularly distinguishable from those due to other azimuths.

In addition to the φ calculations performed for the VAS stimuli, φ values were calculated between BCP patterns of word and nonword stimuli. As speech consists of combinations of narrowband, broadband, and harmonic acoustic signals, these comparisons may be substantially informative.

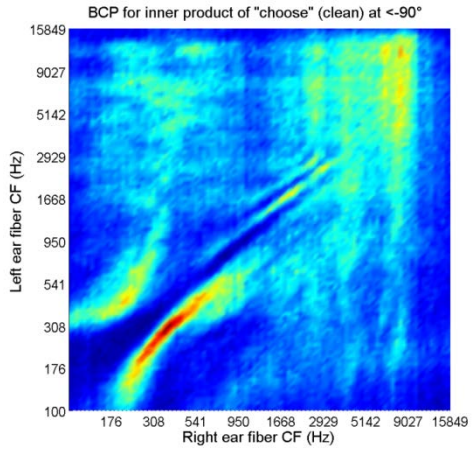
Finally, a model was constructed in order to predict average absolute lateralization errors from $\delta_b$ for a given stimulus:

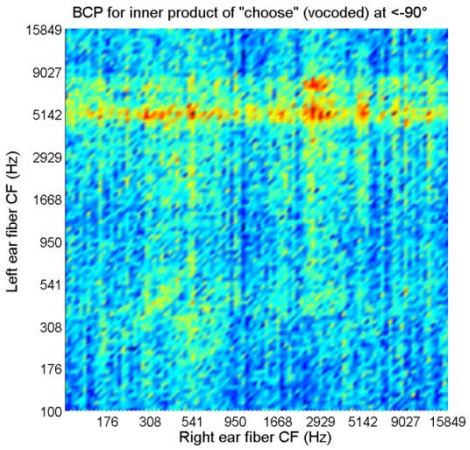
$$RMS\ Error = \beta \times \delta_b + constant$$

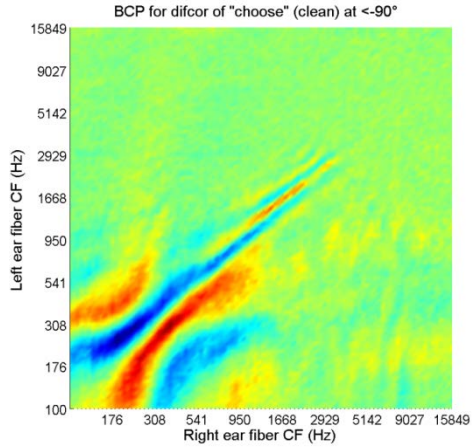
**Results**

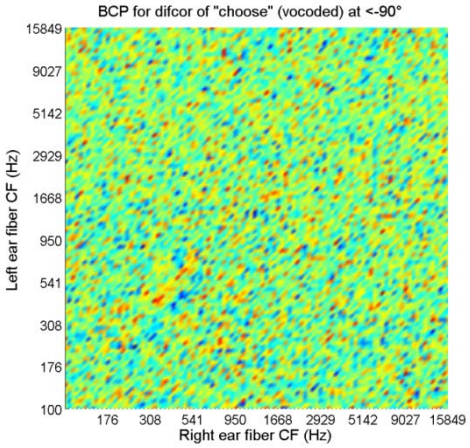
Results varied greatly as functions of stimulus and condition, and graphical depictions will be instrumental in the illustration of these results. Accordingly, Figures 1-3 show the BCPs for the inner product, sumcor, and difcor of a white noise token at 90° azimuth (with full HRTF cues), a clean VAS word (“choose”) at -90°, and a vocoded VAS word (“choose”) at -90°. The color scale used represents large values as red and small values as blue. As explained in the methods section, the color axis is not identical across all figures, as this would wash-out some of the interesting details in some figures, and is not necessary due to the use of normalization.

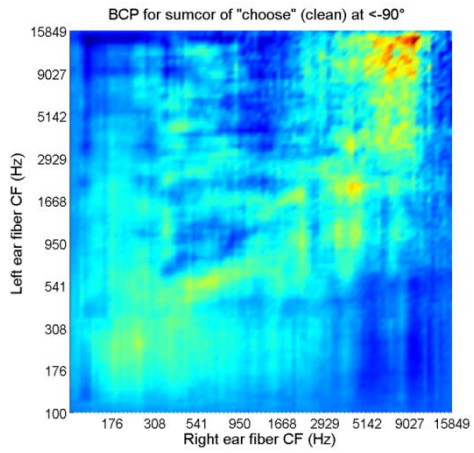
Immediately apparent in the BCPs of the clean word and white noise is the strongly diagonal nature in the streak patterns of highest correlation. This is not unexpected, as signals on the left and right AN fibers should be most strongly correlated when the fibers’ CFs are similar, and a ITD will largely only result in a shift of this pattern to the left or right of the central $CF_{left} = CF_{right}$ diagonal. For the clean VAS stimuli at -90°, we observe the highest correlation values below this central diagonal. This is



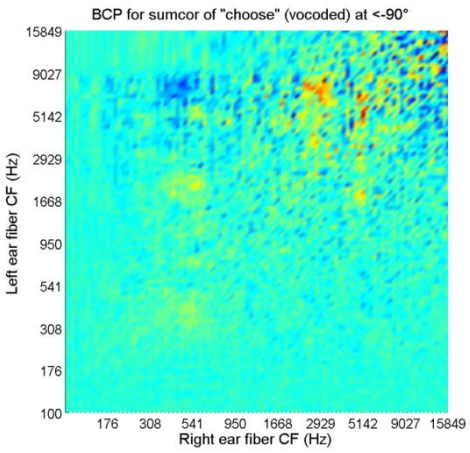



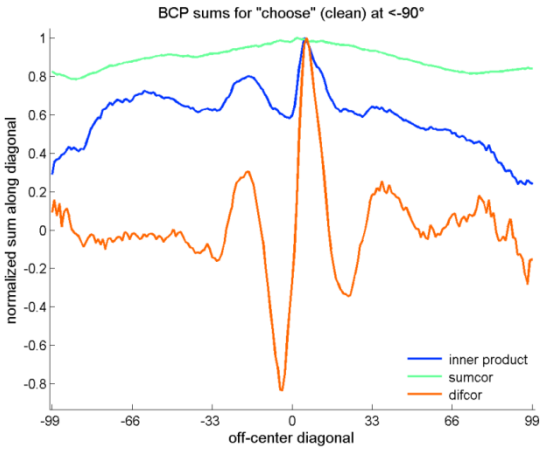

Figure 2. Binaural correlation planes due to inner product, difcor, and sumcor calculations for the word (unprocessed) “choose” at -90°.

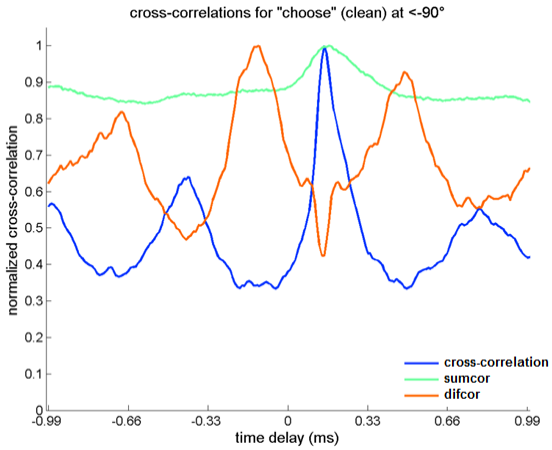
Figure 3. Binaural correlation planes due to inner product, difcor, and sumcor calculations for the vocoded word “choose” at -90°.

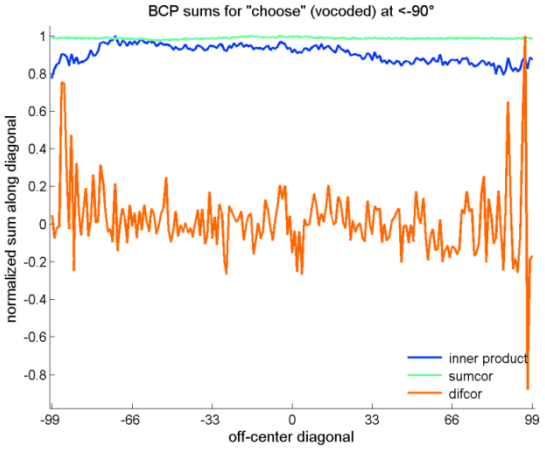
explained by the cochlear delay; signals arrive in the left ear first, and because transduction by low-CF fibers is delayed relative to high-CF fibers, left-ear low-CF fibers will fire in temporal coincidence with higher-CF fibers in the right ear. Thus we see higher-CF fibers in the right ear correlate strongly with relatively lower-CF fibers in the left ear.

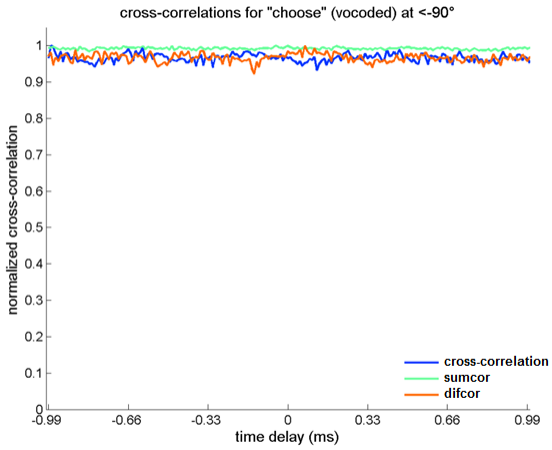
Distinctive patterns are clear in both high- and low-CF regions of the inner product BCPs, although these patterns are different. For lower CFs, the inner product BCPs for clean VAS word and nonword stimuli exhibit the above-mentioned diagonal streaks of correlation. The multiplicity of these streaks is explained by the driving periodic nature of sound for narrowband signals and the ringing response at CF for broadband signals. At higher CFs, a broader blur of correlation is observed. The reason for this blurring is not immediately apparent, but it likely is effected by the loss of phase-locking and refractory effects limiting the ability of single fibers to exhibit clear ringing at these high frequencies. The difcor and sumcor BCPs exhibit these characteristics of low- and high-CF regimes to a more exaggerated degree. The difcor BCP exhibits large undulations in the low-CF streak patterns, while patterns are flat above roughly 3 kHz. This follows from the fact that the difcor is most sensitive to signals arising from phase-locked response to TFS. In contrast, the sumcor exhibits its most distinctive patterns in the smears of correlation throughout the CF range. None of these patterns are visible in the BCPs for vocoded VAS word stimuli.

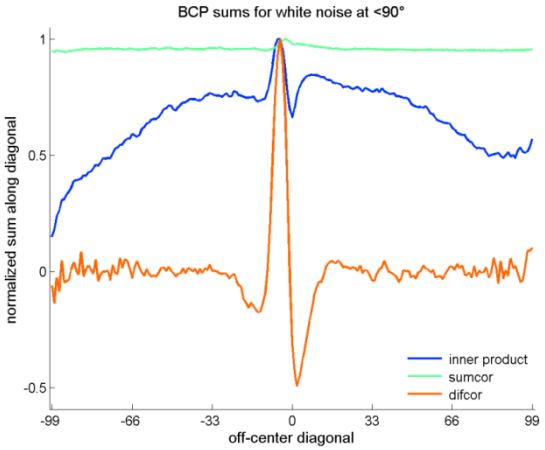
Figure 4 shows the above BCPs' summation, normalized, along the diagonal axes. For the clean VAS word stimulus, the difcor BCP diagonal sum exhibits oscillatory behavior near the central diagonal but then falls to zero, reflecting the fact that the TFS-sensitive signals carried in fibers of disparate CF do not carry related temporal patterns; in fact, owing to their phase-locking properties, they are largely orthogonal. The sumcor BCP diagonal sum, in contrast, displays shallow, monotonic roll-off from its



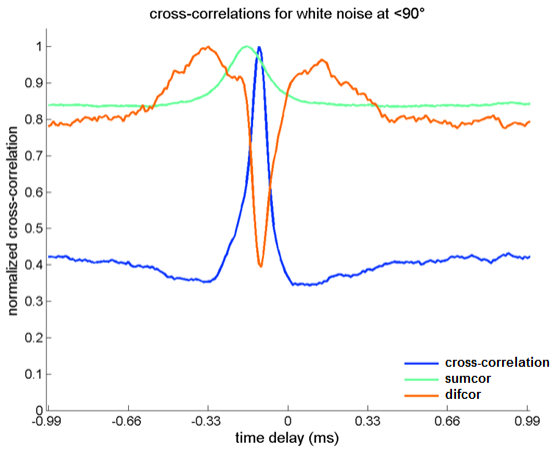




Figure 4. Diagonal sums of the binaural correlation planes due to the inner product, sumcor, and difcor calculations for the clean and vocoded VAS words "choose" and white noise at azimuth -90°. Metrics taken are the off-center diagonal numbers of the highest peaks in these functions.

Figure 5. Sum across all CFs of cross-correlations, sumcors, and difcors of same-CF PSTHs from left and right ears for the clean and vocoded VAS words "choose" and white noise at azimuth -90°. Metrics taken extracted are the time delays of the highest peaks in these functions.

peak, and stays far from zero along all diagonals. The inner product BCP diagonal sum displays a mix of these characteristics- an oscillation near the central diagonal imposed upon a shallow roll-off, staying nonzero for disparate CF pair diagonals. The BCP diagonal sums for white noise display similar patterns, while those for the vocoded VAS word stimulus reflect the nebulousness of the source BCPs.

Figure 5 shows the normalized benchmark metric functions, same-CF cross-correlation functions summed over all 100 CFs, near zero delay. These are characterized by distinct patterns in the cross-correlation, sumcor, and difcor plots for clean VAS word and white noise stimuli, and apparently formless clutter for the vocoded VAS word stimulus. It is interesting to compare these with the BCP diagonal summation plots in Figure 4. Knowing that the extracted metrics are the abscissae of the highest peaks in all of these plots, it is curious that the benchmark metrics are not always indicated by consistent “best” time delays among the co-plotted functions (for example, see the disparate peak locations in Figure 5). In contrast, the BCP diagonal sums tend to have more co-aligned peaks among their three co-plotted measures.

The abscissae of the peaks in the BCP diagonal sums and benchmark functions were extracted as metrics and were found to be generally dependent upon generating azimuth. (As the azimuth-dependent peaks for the BCP diagonal sums stayed relatively near the center diagonal and large excursions were observed near the edges due to low statistics, the location of the peak picked was restricted to ±60 diagonals off the center.) Correlation coefficients due to the linear models between metrics and generating azimuths are shown in Tables 2A and 2B for high and medium SR fibers, respectively. (The model coefficients themselves, β, are not particularly instructive.) The cell colors reflect the correlation coefficients- green for positive values, red for negative values, and yellow for values near zero.

| | ILD | x.XC | x.dif | x.sum | s.IP | s.dif | s.sum | |
|---|---|---|---|---|---|---|---|---|
| click 1 | -0.99 | 1 | 0.45 | 1 | 0.97 | -0.17 | 0.98 | full HRTF cues |
| click 2 | -0.97 | 1 | -0.46 | 1 | 0.22 | -0.36 | 0.98 | |
| click 3 | -0.97 | 0.99 | -0.17 | 1 | 0.24 | 0.07 | 0.98 | |
| noise | 0.99 | 1 | 1 | 0.99 | 0.99 | 0.99 | 0.99 | |
| H0 | 0.98 | 0.99 | 1 | 0.99 | 1 | 1 | 0.97 | |
| H360 | 0.8 | 0.99 | 0.99 | 0.99 | 0.97 | 0.99 | -0.14 | |
| 100Hz | -0.99 | 0.98 | 0.97 | -0.33 | 0.96 | 0.97 | 0.76 | |
| 200Hz | 0.98 | -0.76 | -0.78 | -0.52 | 0.98 | 0.98 | -0.04 | |
| 500Hz | 0.98 | 0.43 | 0.44 | -0.1 | -0.88 | -0.9 | -0.47 | |
| 1.5kHz | 0.85 | 0.37 | 0.37 | 0.26 | -0.33 | -0.23 | -0.25 | |
| 3kHz | 0.96 | 0.17 | 0.29 | 0.45 | -0.4 | -0.16 | -0.59 | |
| 5kHz | 0.99 | 0.23 | 0.08 | 0.27 | -0.96 | 0.07 | 0.14 | |
| click 1 | 0.8 | 1 | 0.11 | 1 | 0.98 | 0.35 | 0.99 | ITD cues only |
| click 2 | 0.9 | 1 | -0.23 | 1 | -0.13 | 0.27 | 0.99 | |
| click 3 | 0.92 | 1 | -0.23 | 1 | 0.39 | -0.24 | 0.99 | |
| noise | -0.9 | 1 | 1 | 1 | 0.99 | 0.99 | 0.96 | |
| H0 | -0.77 | 1 | 1 | 1 | 1 | 1 | 1 | |
| H360 | -0.46 | 1 | 1 | 1 | 0.97 | 0.97 | 0.23 | |
| 100Hz | -0.69 | 1 | 0.99 | -0.42 | 0.97 | 0.98 | 0.69 | |
| 200Hz | -0.66 | -0.62 | -0.71 | -0.19 | 0.99 | 0.99 | 0.97 | |
| 500Hz | -0.85 | 0.09 | 0.39 | -0.15 | 0.74 | -0.7 | -0.31 | |
| 1.5kHz | -0.88 | 0.24 | 0.38 | 0.05 | -0.06 | -0.06 | -0.13 | |
| 3kHz | -0.57 | 0 | 0.03 | -0.26 | -0.18 | -0.09 | -0.02 | |
| 5kHz | -0.96 | -0.23 | 0.23 | 0.29 | 0.15 | -0.48 | 0.04 | |
| click 1 | -0.99 | 0.96 | 0.09 | 0.98 | 0.93 | 0.34 | 0.96 | ILD cues only |
| click 2 | -0.98 | 0.96 | 0.36 | 0.97 | 0.33 | 0.03 | 0.98 | |
| click 3 | -0.98 | 0.95 | 0.44 | 0.97 | 0.94 | -0.09 | 0.98 | |
| noise | 0.99 | 0.59 | 0.47 | 0.93 | 0.65 | 0.63 | 0.69 | |
| H0 | 0.98 | -0.85 | -0.77 | -0.97 | 0.11 | -0.57 | -0.9 | |
| H360 | 0.81 | 0.71 | 0.69 | 0.71 | 0.72 | 0.73 | 0.69 | |
| 100Hz | -0.99 | -0.02 | 0.08 | -0.24 | -0.28 | -0.31 | 0.12 | |
| 200Hz | 0.98 | -0.52 | -0.33 | -0.34 | 0.74 | 0.75 | 0.58 | |
| 500Hz | 0.98 | -0.09 | -0.06 | -0.08 | 0.73 | 0.73 | 0.67 | |
| 1.5kHz | 0.85 | 0.2 | 0.22 | 0.27 | 0.38 | 0.32 | -0.32 | |
| 3kHz | 0.96 | 0 | -0.17 | 0.05 | -0.52 | -0.4 | -0.55 | |
| 5kHz | 0.99 | 0.05 | -0.2 | 0.33 | -0.97 | 0.03 | -0.45 | |

Table 2A. Correlation coefficients between metrics and generating azimuths for nonword stimuli. Measures starting with "x." are drawn from the abscissa (time delay) of the peak in the sum over all 100 CFs of functions (cross-correlation, difcor, and sumcor) calculated from PSTHs for left and right fibers of identical CF. Those starting with "s." (for "stereausis") are drawn from the diagonal number (off-center) of the peak in the diagonal sum of the BCP. These metrics were calculated using simulation data from high SR fibers and 37 azimuths from -90° to +90° in the front horizontal plane. Virtual location cues were generated from the KEMAR HRTF.

| | ILD | x.XC | x.dif | x.sum | s.IP | s.dif | s.sum | |
|---|---|---|---|---|---|---|---|---|
| click 1 | 0.96 | 0.98 | 0.28 | 0.98 | 0.17 | 0.12 | 0.46 | full HRTF cues |
| click 2 | 0.95 | 0.35 | 0.06 | 0.87 | -0.28 | 0.21 | 0.34 | |
| click 3 | 0.95 | 0.38 | -0.23 | 0.79 | 0.02 | 0.13 | 0.35 | |
| noise | 0.98 | 1.00 | 1.00 | 1.00 | -0.48 | 0.99 | 0.96 | |
| H0 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 0.81 | 1.00 | |
| H360 | 0.97 | 1.00 | 1.00 | 1.00 | 0.99 | 0.99 | -0.14 | |
| 100Hz | -0.99 | 0.74 | 0.77 | -0.54 | 0.92 | 0.69 | -0.32 | |
| 200Hz | 0.98 | -0.51 | -0.65 | 0.06 | 0.98 | 0.94 | 0.87 | |
| 500Hz | 0.98 | 0.18 | 0.11 | -0.13 | 0.99 | -0.86 | -0.52 | |
| 1.5kHz | 0.88 | 0.15 | -0.01 | 0.26 | 0.17 | -0.20 | -0.58 | |
| 3kHz | 0.96 | 0.04 | -0.01 | 0.08 | -0.07 | 0.26 | -0.01 | |
| 5kHz | 0.99 | 0.30 | -0.11 | 0.26 | -0.91 | 0.35 | 0.49 | |
| click 1 | -0.96 | 0.98 | 0.07 | 0.94 | 0.93 | 0.42 | 0.83 | ITD cues only |
| click 2 | -0.85 | -0.05 | -0.25 | -0.07 | 0.11 | 0.11 | -0.06 | |
| click 3 | -0.85 | | 0.11 | 0.04 | 0.04 | 0.30 | 0.17 | |
| noise | -0.68 | 1.00 | 1.00 | 1.00 | 0.83 | 0.99 | 0.88 | |
| H0 | -0.20 | 1.00 | 1.00 | 1.00 | 0.98 | 1.00 | 1.00 | |
| H360 | -0.30 | 1.00 | 1.00 | 1.00 | 0.95 | 0.97 | -0.14 | |
| 100Hz | -0.54 | 0.80 | 0.90 | -0.42 | 0.86 | 0.82 | 0.83 | |
| 200Hz | -0.62 | -0.59 | -0.32 | 0.30 | 0.99 | 0.96 | 0.94 | |
| 500Hz | -0.45 | 0.46 | 0.37 | 0.18 | 0.99 | -0.48 | 0.39 | |
| 1.5kHz | -0.39 | 0.16 | 0.07 | 0.07 | -0.17 | -0.17 | 0.13 | |
| 3kHz | -0.51 | -0.12 | 0.27 | 0.03 | -0.15 | -0.18 | -0.16 | |
| 5kHz | -0.89 | | 0.13 | -0.11 | 0.18 | 0.25 | -0.17 | |
| click 1 | 0.96 | 0.63 | 0.08 | 0.93 | 0.63 | 0.20 | 0.59 | ILD cues only |
| click 2 | 0.94 | -0.17 | 0.03 | -0.03 | -0.34 | 0.04 | -0.50 | |
| click 3 | 0.94 | -0.26 | -0.11 | 0.16 | -0.47 | -0.16 | -0.58 | |
| noise | 0.98 | 0.51 | 0.40 | 0.91 | -0.28 | 0.55 | | |
| H0 | 0.99 | -0.86 | -0.58 | -0.64 | -0.29 | -0.25 | -0.66 | |
| H360 | 0.97 | 0.60 | 0.48 | 0.85 | 0.57 | 0.58 | 0.63 | |
| 100Hz | -0.99 | -0.19 | 0.10 | -0.16 | 0.10 | 0.15 | 0.28 | |
| 200Hz | 0.98 | -0.12 | -0.09 | -0.21 | 0.76 | 0.59 | 0.48 | |
| 500Hz | 0.98 | -0.03 | 0.06 | 0.15 | 0.65 | 0.64 | 0.54 | |
| 1.5kHz | 0.88 | -0.23 | 0.06 | -0.20 | 0.25 | 0.36 | -0.35 | |
| 3kHz | 0.96 | -0.07 | -0.10 | 0.09 | -0.27 | 0.21 | -0.12 | |
| 5kHz | 0.99 | 0.11 | -0.26 | 0.01 | -0.92 | 0.61 | 0.48 | |

Table 2B. Correlation coefficients between metrics and generating azimuths for nonword stimuli using simulation data from medium SR fibers. Blank entries indicate where calculations were not available.

It is interesting to examine in these tables how the measures change between the two SRs. Especially striking is how calculated ILDs are precise cues for clicks with medium spontaneous rate fibers, but how they produce *negative* correlations with high SR fibers. The calculated ILD is otherwise a metric strongly correlated to generating azimuth for stimuli with full HRTF cues and ILD cues only, the only exception being a 100 Hz tone, for which the ILD gives a strongly anticorrelated cue. The reason for this is likely twofold. First, the physical ILD at 100 Hz is negligible. Second, a slightly louder signal in the ipsilateral ear may cause more AN fibers to be in states of refraction, limiting the number of spikes generated on that side. When only ITD cues were present, the neural ILD did not provide a positively-correlated metric. When only ILD cues were present, the calculated ILD generally provided the best cues. A strange finding is the difference in predictive power of timing metrics for azimuths of harmonic complexes of sine and random phase when only ILD cues were present. For both fiber SRs, sine-phase harmonic complex ITD metrics were anticorrelated with azimuth while random-phase harmonic complex ITD metrics were positively correlated with azimuth. This finding in itself is supremely curious and warrants further investigation.

The benchmark and BCP diagonal sum (stereausis) ITD metrics provided comparably strong cues to azimuth for most conditions. An exception is the prediction of azimuth for a 200 Hz tone, which the benchmarks metric do poorly, while the stereausis metrics do fairly. Timing metrics are universally poor at predicting generating azimuth for high-frequency (≥ 1.5kHz) narrowband tones.

Correlation coefficients between metrics and generating azimuths and those between metrics and *response azimuths* for clean and vocoded VAS stimuli are shown in Tables 3A and 3B, respectively. For both sets of correlation coefficients, neural ILDs were found to be strongly predictive of azimuth, benchmark timing measures were found to be strongly predictive of azimuth for clean VAS word stimuli,

| | ILD | x.XC | x.dif | x.sum | s.IP | s.dif | s.sum | |
|---|---|---|---|---|---|---|---|---|
| beam | 0.93 | 0.98 | 0.99 | 0.99 | 0.95 | 0.95 | 0.44 | clean |
| cape | 0.94 | 0.97 | 0.97 | 0.98 | 0.95 | 0.62 | -0.02 | |
| car | 0.93 | 0.99 | 1.00 | 0.99 | 0.99 | 0.98 | 0.74 | |
| choose | 0.95 | 0.96 | 0.97 | 0.99 | 0.92 | 0.95 | 0.35 | |
| chore | 0.94 | 0.98 | 0.98 | 0.99 | 0.98 | 0.98 | 0.67 | |
| beam | 0.86 | 0.00 | 0.19 | 0.01 | -0.71 | -0.02 | -0.61 | vocoded |
| cape | 0.93 | 0.44 | 0.35 | 0.39 | -0.69 | -0.25 | -0.48 | |
| car | 0.89 | -0.04 | 0.03 | 0.46 | -0.75 | 0.23 | -0.11 | |
| choose | 0.91 | 0.00 | -0.06 | 0.26 | -0.78 | -0.52 | -0.55 | |
| chore | 0.89 | -0.04 | -0.06 | 0.25 | -0.22 | -0.47 | -0.68 | |

Table 3A. Correlation coefficients between metrics and generating azimuths for clean and vocoded VAS stimuli using simulation data from high SR fibers and 19 azimuths from -90° to +90° in the front horizontal plane. Virtual location cues were generated from the listener's recorded HRTF.

| | ILD | x.XC | x.dif | x.sum | s.IP | s.dif | s.sum | |
|---|---|---|---|---|---|---|---|---|
| beam | 0.93 | 0.98 | 0.98 | 0.99 | 0.96 | 0.96 | 0.93 | clean |
| cape | 0.94 | 0.98 | 0.97 | 0.98 | 0.96 | 0.60 | -0.17 | |
| car | 0.93 | 0.98 | 0.99 | 0.98 | 0.98 | 0.97 | 0.74 | |
| choose | 0.96 | 0.96 | 0.97 | 0.99 | 0.92 | 0.95 | 0.47 | |
| chore | 0.94 | 0.99 | 0.98 | 0.99 | 0.98 | 0.98 | 0.69 | |
| beam | 0.90 | -0.20 | -0.04 | -0.16 | -0.66 | -0.15 | -0.44 | vocoded |
| cape | 0.95 | 0.66 | 0.51 | 0.42 | -0.62 | -0.11 | -0.61 | |
| car | 0.94 | -0.04 | -0.25 | 0.46 | -0.71 | 0.28 | -0.12 | |
| choose | 0.92 | -0.25 | -0.10 | 0.18 | -0.73 | -0.29 | -0.58 | |
| chore | 0.92 | -0.11 | -0.04 | -0.06 | -0.30 | -0.61 | -0.60 | |

Table 3B. Correlation coefficients between metrics and average listener response azimuths for clean and vocoded VAS stimuli.

| | ILD | x.XC | x.dif | x.sum | s.IP | s.dif | s.sum |
|---|---|---|---|---|---|---|---|
| clean | 0.003 | 0.001 | -0.002 | -0.001 | 0.001 | -0.003 | 0.095 |
| vocoded | 0.033 | -0.057 | -0.075 | -0.105 | 0.023 | 0.028 | 0.013 |

Table 3C. Average difference between correlation coefficients due to metric vs azimuth and metric vs response. This may indicate which cues the listener was actually using for lateralization.

| | ILD | x.XC | x.dif | x.sum | s.IP | s.dif | s.sum |
|---|---|---|---|---|---|---|---|
| clean | 17° | 8° | 8° | 7° | 13° | 24° | 39° |
| vocoded | 34° | 229° | 254° | 150° | 276° | 161° | 62° |

Table 3D. Average absolute errors between model predictions and listener response.

stereausis timing measures' predictive abilities were mixed for clean VAS word stimuli, and all timing measures were poor at predicting vocoded VAS word stimuli azimuths. Table 3C shows how much better, on average, the metrics predict *response* azimuths than *generating* azimuths. These differences in correlation coefficients are very small, but may suggest that the listener's response relies in part on stereausis sumcor cues for clean VAS words and ILDs for vocoded VAS words. Table 3D shows the average absolute differences between model azimuth predictions based on the metrics and listener responses. These values are larger than the average RMS lateralization errors made by the listener, indicating that none of these metrics alone can account for lateralization performance.

Figure 6 shows the model predictions for response azimuth with clean and vocoded VAS based on each of the predictors. The model coefficients here were derived from linear best fits for clean VAS metrics and listener response. Neither these nor the correlation coefficients differed significantly from

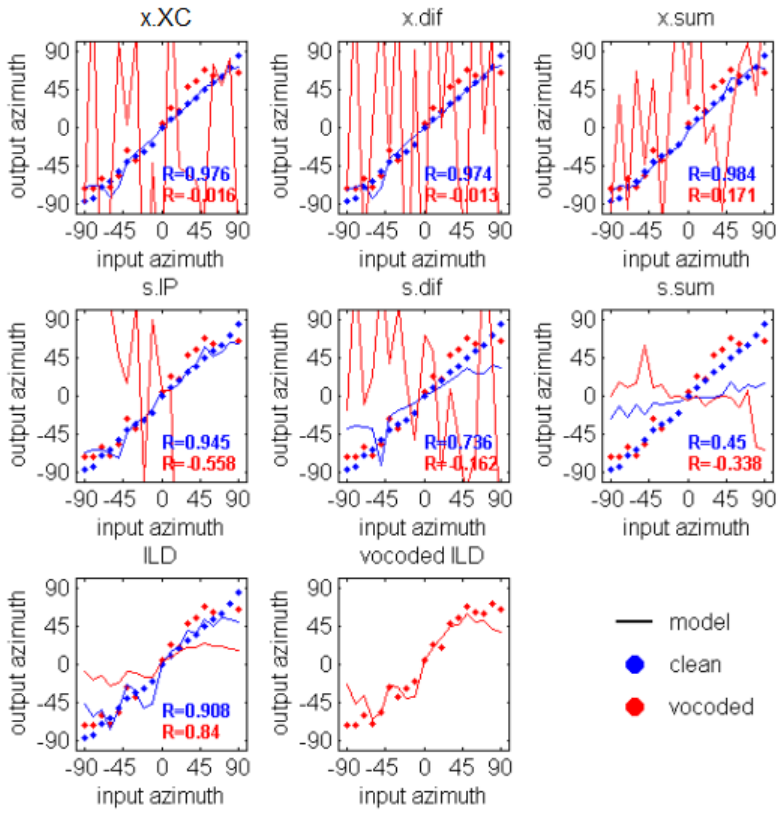



Figure 6. Model responses and correlation coefficients for clean and vocoded VAS stimuli using seven metrics as predictors. Dots indicate average listener responses for each virtual input azimuth and solid lines depict model predictions. Model coefficients were calculated from least-squares fits of metrics and response data from clean VAS stimuli. Models calculated using generating azimuths instead of listener responses resulted in identical correlation coefficients up to three decimal places. The eighth panel shows the linear model of ILD prediction of vocoded VAS response azimuth generated from that data.

| | ILD | x.XC | x.dif | x.sum | s.IP | s.dif | s.sum | |
|---|---|---|---|---|---|---|---|---|
| click 1 | 3644° | 12° | 35° | 9° | 35° | 47° | 40° | clean |
| click 2 | 3249° | 15° | 50° | 11° | 45° | 48° | 40° | |
| click 3 | 3253° | 16° | 47° | 11° | 45° | 46° | 40° | |
| noise | 86° | 8° | 9° | 8° | 14° | 17° | 49° | |
| H0 | 100° | 14° | 10° | 44° | 28° | 29° | 39° | |
| H360 | 724° | 9° | 9° | 9° | 18° | 22° | 48° | |
| 100Hz | 1865° | 8° | 9° | 50° | 91° | 113° | 47° | |
| 200Hz | 820° | 51° | 52° | 52° | 27° | 30° | 46° | |
| 500Hz | 29° | 43° | 43° | 47° | 51° | 52° | 51° | |
| 1500Hz | 17° | 43° | 44° | 44° | 54° | 53° | 52° | |
| 3kHz | 37° | 45° | 44° | 40° | 64° | 56° | 59° | |
| 5kHz | 24° | 44° | 46° | 43° | 73° | 46° | 45° | |
| click 1 | 1394° | 182° | 75° | 130° | 101° | 48° | 57° | vocoded |
| click 2 | 1246° | 165° | 51° | 118° | 55° | 47° | 57° | |
| click 3 | 1248° | 161° | 49° | 119° | 56° | 49° | 57° | |
| noise | 25° | 229° | 258° | 134° | 287° | 283° | 131° | |
| H0 | 27° | 298° | 306° | 295° | 137° | 129° | 66° | |
| H360 | 242° | 214° | 242° | 159° | 195° | 174° | 47° | |
| 100Hz | 728° | 225° | 240° | 51° | 795° | 834° | 121° | |
| 200Hz | 279° | 53° | 55° | 53° | 419° | 393° | 48° | |
| 500Hz | 27° | 50° | 51° | 49° | 37° | 51° | 45° | |
| 1500Hz | 35° | 50° | 50° | 47° | 46° | 57° | 46° | |
| 3kHz | 26° | 49° | 49° | 48° | 95° | 69° | 50° | |
| 5kHz | 29° | 49° | 48° | 47° | 146° | 49° | 50° | |

Table 4. Average absolute errors in predictions of response azimuths (vs (vs listener response) to clean and vocoded VAS stimuli using model coefficients derived from nonword stimuli.

those derived from clean VAS metrics and generating azimuths, indicating the listener performed the lateralization task nearly optimally for these clean VAS stimuli. Only the calculated ILD metric could partially predict the listener's response azimuth for vocoded VAS stimuli, but even here, the model does not do as well as the listener. The bottom-center panel of Figure 6 shows a model whose coefficient was determined from vocoded VAS stimuli and listener responses. Close examination of this panel indicates that the pattern of listener response is closely followed for all except the most lateral azimuths.

The metric coefficients derived from modeling of generating azimuth of nonword stimuli were applied to the metrics of VAS word stimuli. Data used were from high-SR simulations with full HRTF cues. Table 4 shows the average absolute differences between the resulting predicted azimuths and

listener responses. The low errors for clean VAS word timing metrics suggest similar patterns in the BCPs for clean VAS words and nonword stimuli. Relatively low errors for azimuth prediction of vocoded VAS words using ILD metrics for some nonwords' metric coefficients indicates the prevalence of this cue.

The cos(φ) for all pairwise azimuth combinations between BCPs for clean VAS word and nonword stimuli, averaged across the five words, are shown in Figures 7-9 for inner product, sumcor, and difcor BCPs. As with the BCPs themselves, these plots use the heat map convention, with the lowest values on each plot represented by blue and the highest values by red. These figures show the azimuth dependence of similarity in BCPs for word and nonword stimuli. It might be expected that BCPs of identical calculation and azimuth might be highly similar, but this is only occasionally the case. There are

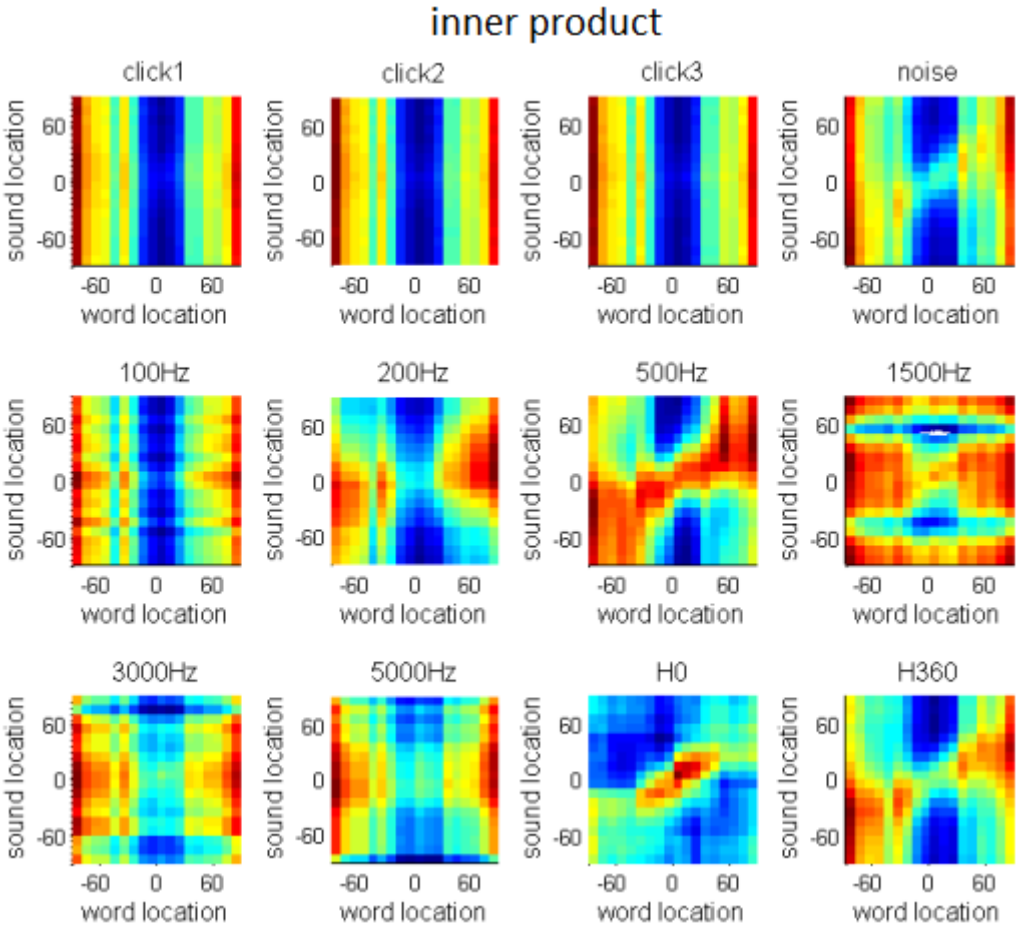
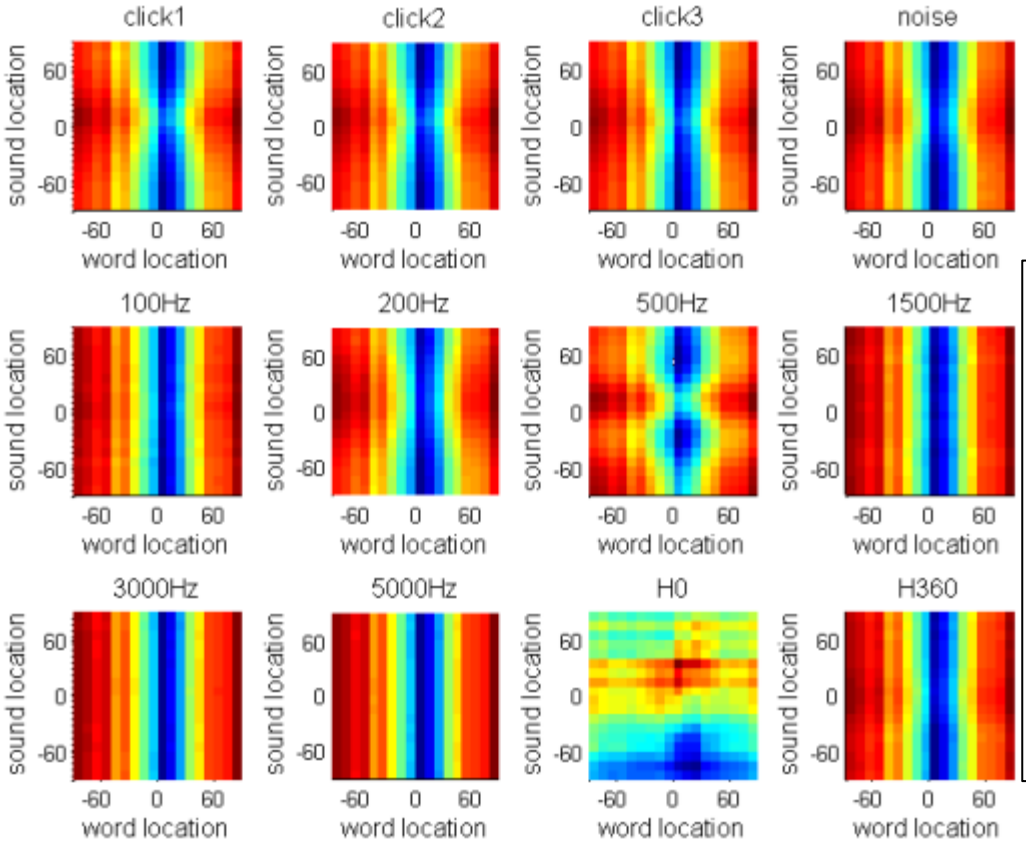
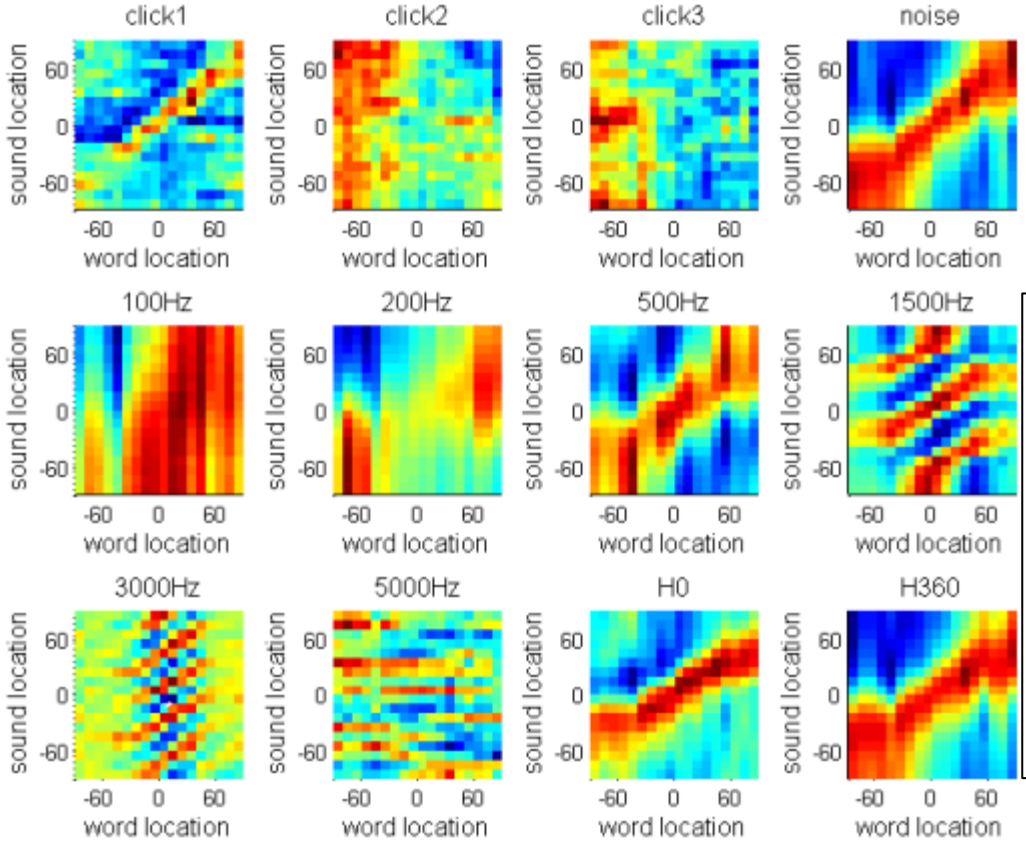



Figure 7. Plots of cos(φ) between inner product BCPs for each nonword stimulus and the average of inner product BCPs for all five clean VAS word stimuli as functions of azimuth. As with the BCP heat maps, the colors represent the spectrum of minimum to maximum values on each plot from blue to red.




sumcor

Figure 8. Plots of cos(φ) between sumcor BCPs for each nonword stimulus and the average of sumcor BCPs for all five clean VAS word stimuli as functions of azimuth




difcor

Figure 9. Plots of cos(φ) between difcor BCPs for each nonword stimulus and the average of difcor BCPs for all five clean VAS word stimuli as functions of azimuth.

some trends of pattern similarity among BCPs due to clean VAS word and nonword stimuli, and the shape and clarity of these trends may indicate which nonword stimuli are most like the word stimuli. For example, the panel in figure 7 for the 500 Hz tone shows the only strong trend of similar BCP patterns with similar azimuths for the inner product calculation. As seen in figure 9, several nonwords' difcor BCP patterns appear to be similar to the average BCP patterns for words of nearby azimuths. Figure 10 shows the complimentary comparisons, cos(φ) for all pairwise azimuth combinations between BCPs for each clean VAS word and the averages of all 12 nonword stimuli. The images from similar calculations with vocoded VAS stimuli are not as enlightening.

Figure 11 shows the cos(φ) values for all azimuth pair comparisons of BCPs for clean VAS words.

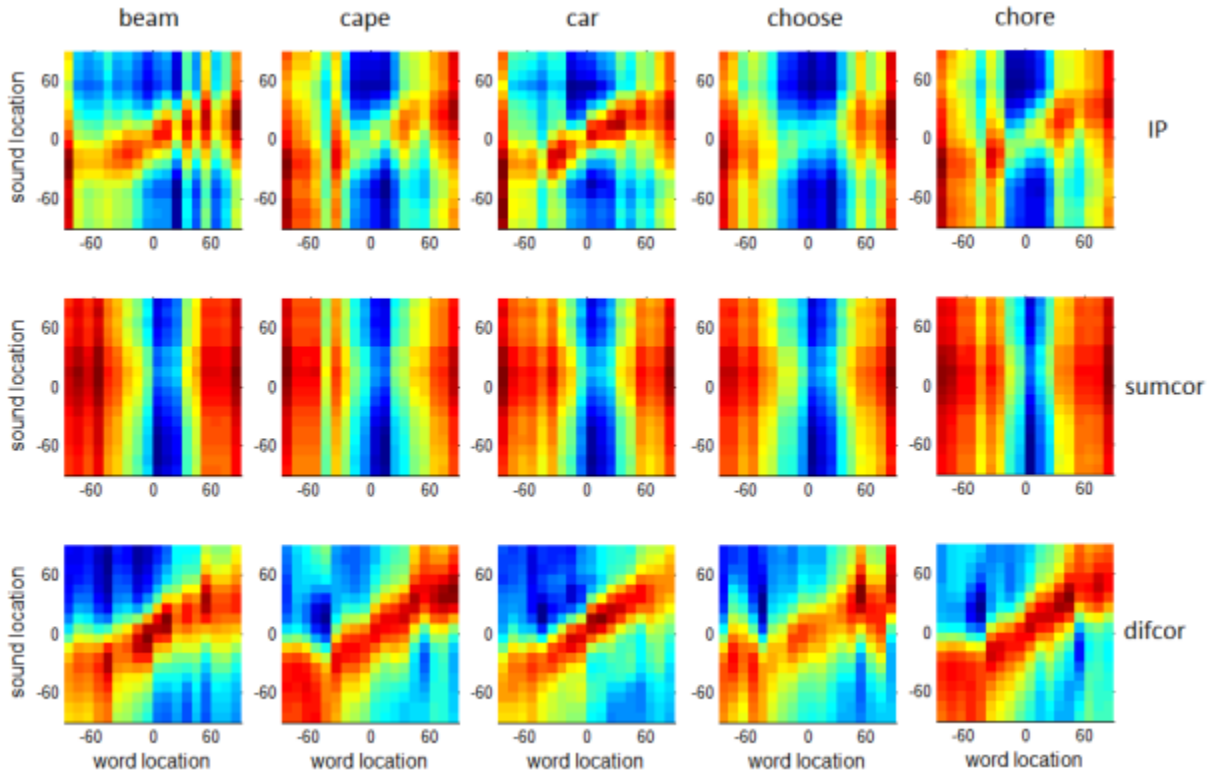



Figure 10. Plots of cos(φ) between average BCPs for all nonword stimuli and BCPs for each clean VAS word stimulus as functions of azimuth. The three rows show cos(φ) azimuth plane plots for BCPs from inner product, sumcor, and difcor calculations from top to bottom.

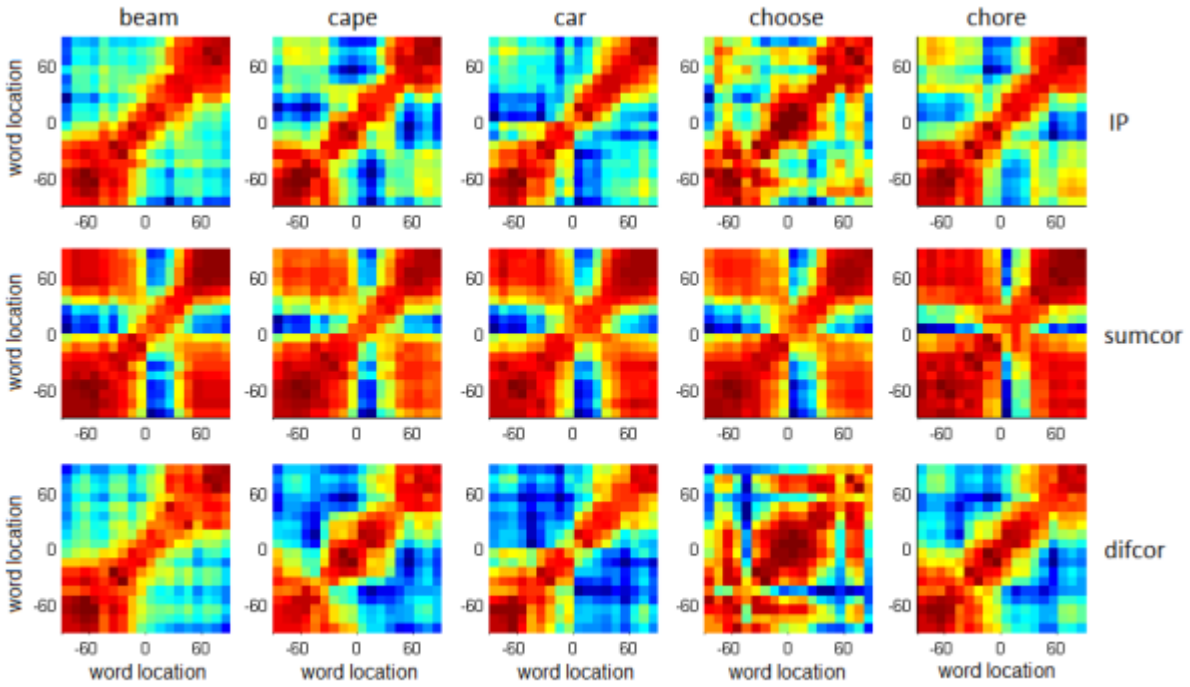



Figure 11. Plots of cos(φ) between BCPs for each clean VAS word stimulus as functions of azimuth. Values on the diagonal are determined from the average of cos(φ) values on adjacent azimuth pairs.

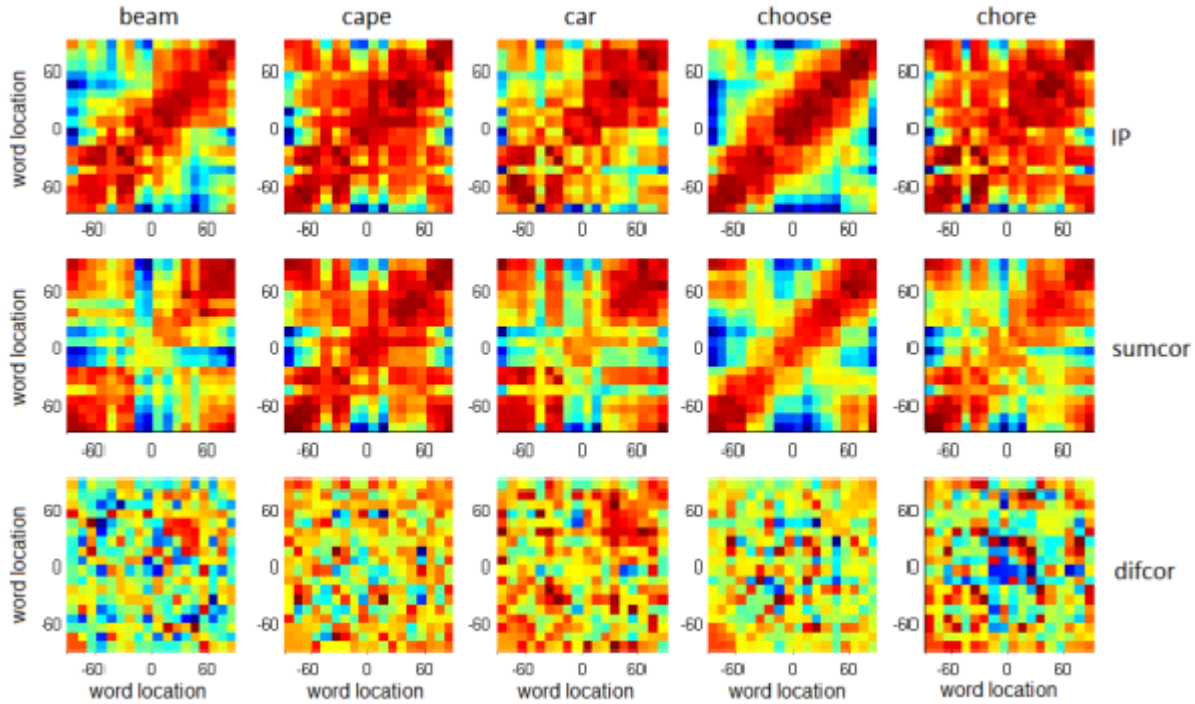



Figure 12. Plots of cos(φ) between BCPs for each vocoded VAS word stimulus as functions of azimuth.

These images show how similar or different the BCPs are between sets of locations for a given clean VAS word and calculation. Because these calculations are comparisons among BCPs of different azimuths for the same stimuli, these plots have had their diagonals "blunted", i.e., the cos(φ) values along the diagonal are replaced by averages of the cos(φ) values on adjacent azimuth pairs. This is done to remove the artificiality of cos(φ) = 1 for identical BCPs, which distorts the color axis of the plot. A thin diagonal stripe would indicate that a given BCP is unique to its azimuth. The wider this diagonal strip, the more similar a given BCP is to those for adjacent azimuths. The reason for the wide "wings" in the sumcor cos(φ) azimuth plane are unknown, but this does indicate that envelope ITDs may not provide unique cues. The narrowest central strips seem to be near the 0° azimuth, indicating that BCPs are most unique around that point. This makes sense because the change in ITD with azimuth is largest near 0°. The same plots are shown for vocoded VAS words in figure 12. It is at once clear that TFS timing cues in the difcor BCPs are useless as unique predictors of azimuth. Some of these cos(φ) azimuth plane plots for inner product and sumcor BCPs comparisons do seem to show diagonal strips, however, indicating that the BCPs do carry marginally unique, azimuth-dependent pattern cues and that these cues are related to the temporal envelopes. The same cos(φ) calculation azimuth planes are shown for nonword stimuli in Figures 13-15. Perhaps most informative from these is the pattern ambiguities in BCPs for 1.5 kHz and 3 kHz tones. Also present are the "wings" in the cos(φ) azimuth plane plots for sumcor BCPs with the clicks and white noise.

How unique a BCP is to a given azimuth is illustrated by how different values along the (blunted) diagonal of these cos(φ) azimuth planes are from all other values on the plane. This is quantified in the blunted uniqueness angle $\delta_b$, which is shown for VAS stimuli and nonword stimuli in Figures 16 and 17, respectively. The higher the value of $\delta_b$, the more dissimilar values along the diagonal are from all others, i.e., the more unique the BCP is for a given azimuth. In the case of BCP pattern template

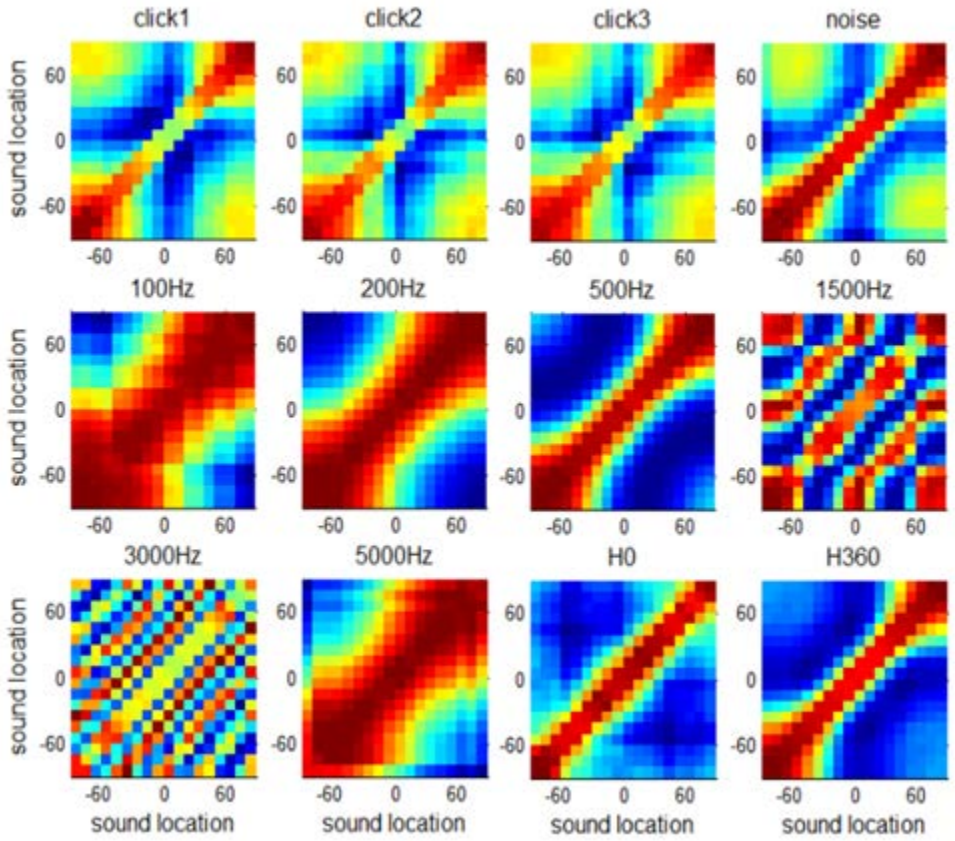



Figure 13. Plots of cos(φ) between inner product BCPs for each stimulus as functions of azimuth.

matching as a mechanism for lateralization abilities, higher $\delta_b$ values would seem to lend to easier lateralization and smaller errors. This trend is certainly the case between clean and vocoded VAS words; clean VAS words produced low lateralization errors and had high $\delta_b$ values, whereas vocoded VAS words produced higher lateralization errors and had low $\delta_b$ values. This trend was modeled using $\delta_b$ values due to inner product, sumcor, and difcor BCPs, and resulting statistics are summarized in table 5. All possible combinations of $\delta_b$ resulted in negative coefficients and strong models, validating the assertion that more unique BCP patterns, indicated by larger $\delta_b$ values, produce lower lateralization errors. The psychoacoustic lateralization errors and those predicted by the model with $\delta_b$ values from inner product, sumcor, and difcor BCPs are shown in figure 18.

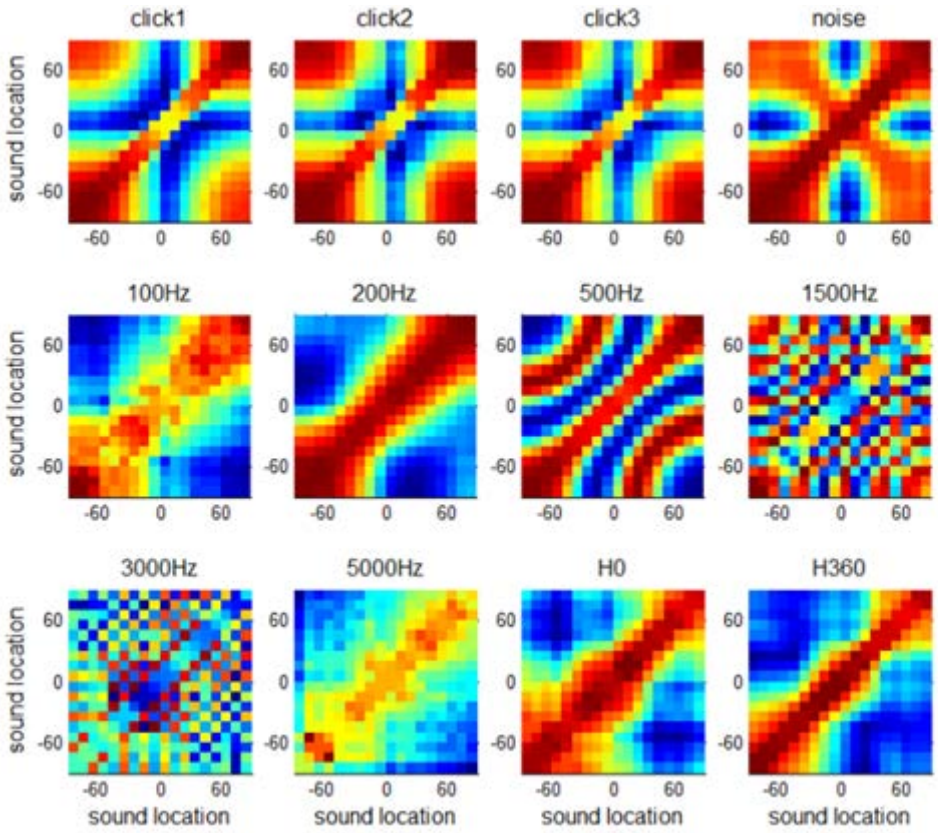



Figure 14. Plots of cos(φ) between sumcor BCPs for each nonword stimulus as functions of azimuth.

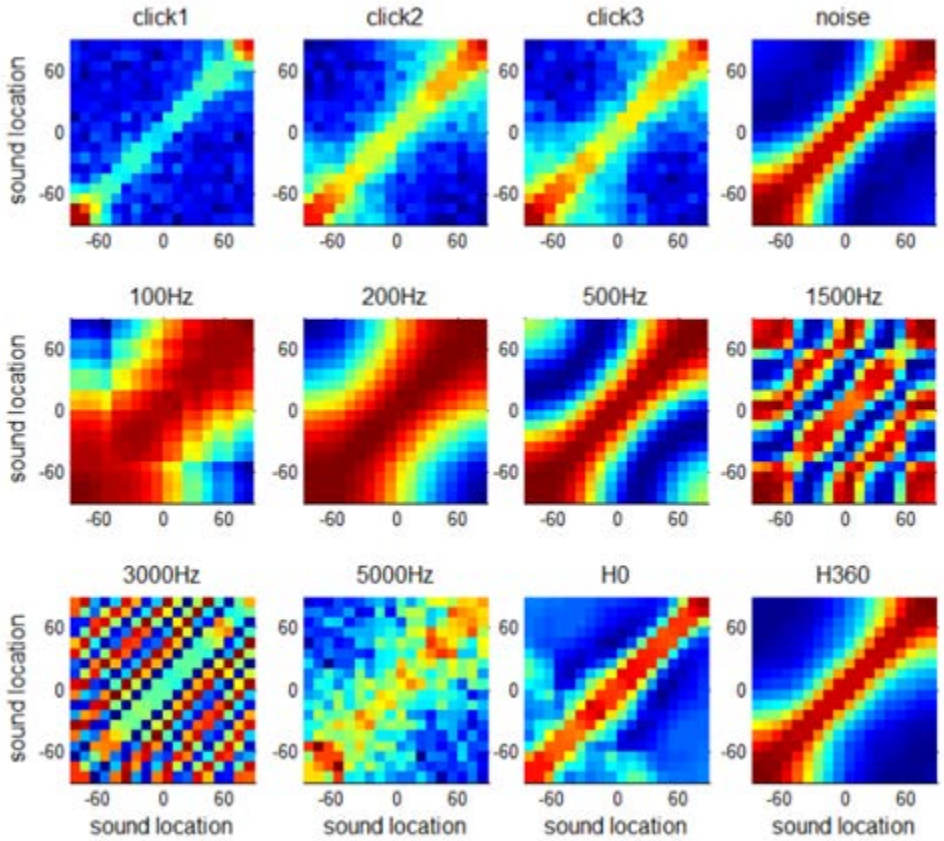



Figure 15. Plots of cos(φ) between difcor BCPs for each nonword stimulus as functions of azimuth.

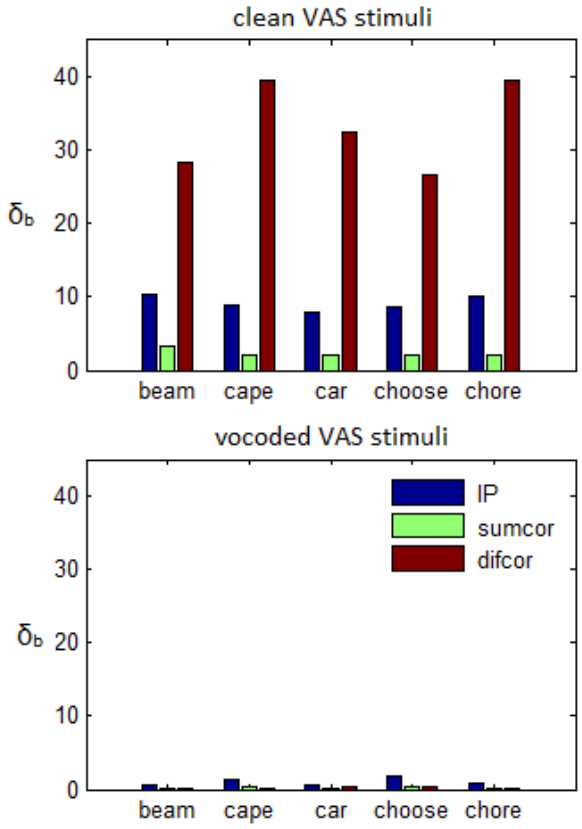



Figure 16. Measures of BCP uniqueness with azimuth, $\delta_b$, for clean and vocoded VAS stimuli.

| variable | β | p | $R^2$ |
|---|---|---|---|
| IP $\delta_b$ | -1.288 | <0.001 | 0.929 |
| sumcor $\delta_b$ | -4.681 | <0.001 | 0.880 |
| difcor $\delta_b$ | -0.305 | <0.001 | 0.871 |
| IP $\delta_b$ | -1.260 | 0.064 | 0.929 |
| sumcor $\delta_b$ | -0.107 | 0.962 | |
| IP $\delta_b$ | -1.182 | 0.047 | 0.930 |
| difcor $\delta_b$ | -0.027 | 0.829 | |
| sumcor $\delta_b$ | -2.579 | 0.063 | 0.924 |
| difcor $\delta_b$ | -0.153 | 0.085 | |
| IP $\delta_b$ | -0.928 | 0.492 | |
| sumcor $\delta_b$ | -0.643 | 0.832 | 0.930 |
| difcor $\delta_b$ | -0.049 | 0.774 | |

Table 5. Model coefficients and their significance for predicting listener localization errors for each word from $d_b$. All seven combinations of $d_b$ resulted in negative coefficients b on all variables.

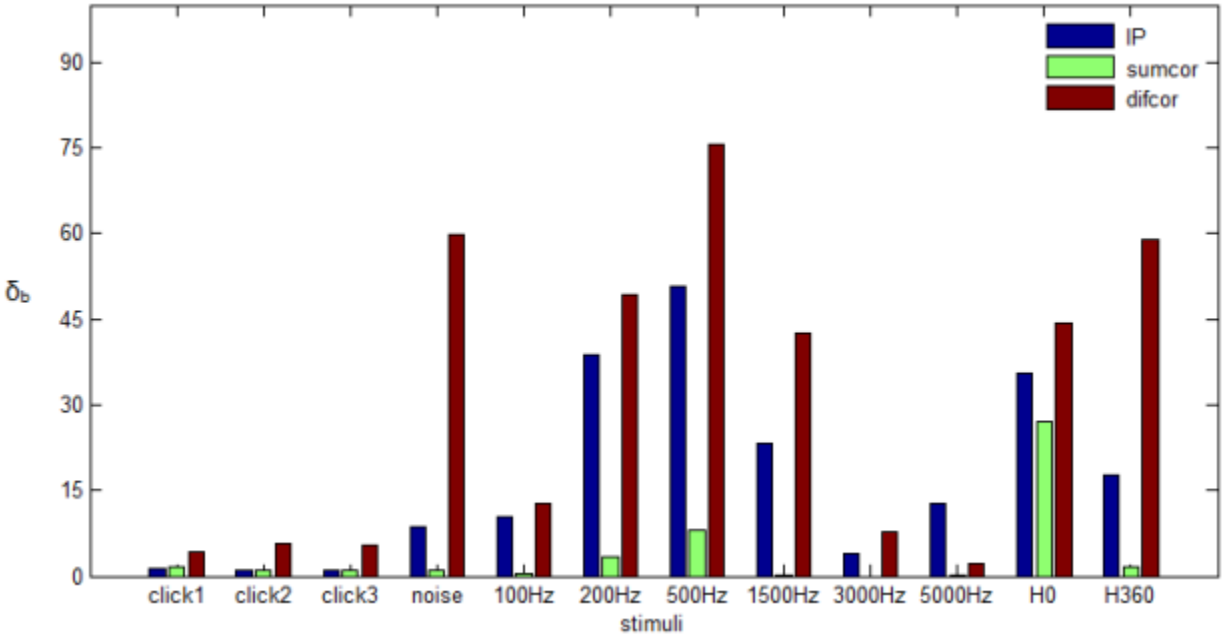



Figure 17. Measures of BCP uniqueness with azimuth, $\delta_b$, for nonword stimuli.

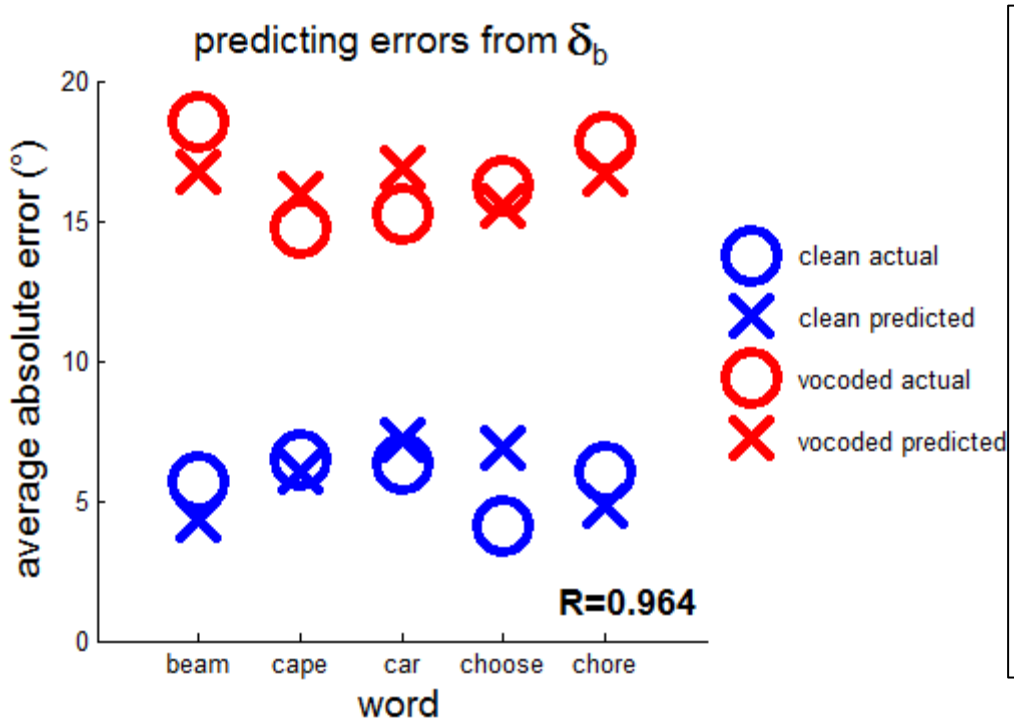



Figure 18. Measured and predicted values of average absolute lateralization errors for each clean and vocoded VAS word. The model was generated using $\delta_b$ values from inner product, sumcor, and difcor BCPs.

**Discussion**

This study used outputs of a computational AN model (Zilany et al., 2009) to calculate seven different azimuth-dependent metrics in order to investigate possible neural mechanisms for the computation of and resulting percept of lateral position. These metrics consisted of the following ILD and ITD calculations: 1) neural ILD calculated from left-right difference in number of simulated spikes generated in all of 100 same-CF fiber pairs; time delay abscissae for peaks in the sum of same-CF 2) cross correlation, 3) difcor, and 4) sumcor functions; diagonal number abscissae for peaks in the diagonal sum of all CF pair BCPs for 5) inner product, 6) difcor, and 7) sumcor calculations. The stimuli consisted of nonword and clean and vocoded VAS word stimuli. Location cues consisted of full HRTF cues, ITDs only, or ILDs only, and were generated from KEMAR or a listener's measured HRTF. The difcor and sumcor metrics are thought to represent neural response due to TFS and envelope, respectively.

Results indicated that the ILD metrics provided a consistently useful cue for lateral position for nonword stimuli when ILD cues were present and for clean and vocoded words. The exceptions to this are the three different clicks when high SR fibers were simulated and the 100 Hz sine tone for both high and medium SR fibers. Possible explanations for these cases were offered in the preceding section. An important factor regarding the clicks may be that they are the only stimuli which was not on-going, i.e., they consisted of positive and/or negative phases in the first 1 or 2 10μs frames of the stimulus, then silence. The majority of the simulated neural response, then, was due to ringing and spontaneous activity.

When ITD cues were present, the time delay associated with the peak in the sum of same-CF cross-correlation functions for high SR fibers provided an excellent cue for lateral position of clicks, white noise, harmonic complexes, 100 Hz tones, and clean words. For narrowband signals above 100 Hz, the metric is a poor cue, possibly due to the increased multiplicity of peaks in the cross-correlation function and the loss of phase-locked response at the higher frequencies. When only ILD cues were available, this metric was good for clicks. This finding may be attributable to the level dependence of the BM motion phase to which ANs respond. With high SR fibers, the location of the peak in the difcor correlogram was a useful cue only when ITDs were present and only for white noise, harmonic complexes, 100 Hz tones, and clean words. The location of the peak in the sumcor correlogram was a useful cue for all broadband stimuli when ITDs were present and for clicks and white noise when only ILDs were present. None of these three metrics could predict the generating azimuth for narrowband signals above 100 Hz.

The location of the peak in the diagonal sum in the inner product BCP, or the “stereausis inner product”, provided a good cue for the azimuth of the monophasic click, white noise, harmonic

complexes, 100 Hz and 200 Hz tones, and clean words when ITDs were present and for high SR fibers. In addition, for medium SR fibers, this metric provided excellent azimuth prediction for the 500 Hz tone when ITDs were present, the only metric to do so. The location of the stereausis difcor peak was a good predictor of azimuth for white noise, harmonic complexes, 100 Hz and 200 Hz tones, and most clean words when ITDs were present. The location of the stereausis sumcor peak was a good predictor of azimuth for clicks, white noise, and the sine phase harmonic complex when ITDs were present with high SR fibers. Again, these stereausis metrics could not predict high-frequency narrowband signals' generating azimuths.

For high SR fibers when only ILDs were present, the location of peaks in the cross correlation and sumcor correlograms and in the inner product and sumcor BCP diagonal sums could predict the azimuth for clicks. Again, this may be due to the level dependence of phase of transduction and that most of the neural signal was due to ringing. It is curious however, that the difcor calculations did not produce useful cues. This indicates that the relative timing of envelope transduction is a function of ILDs. Similarly curious, all six ITD-based metrics produced serviceable azimuth cues for the random-phase harmonic complex when only ILDs were present, but strongly anti-correlated cues for the sine phase harmonic complex. Recognizing the difference between the two signals is purely temporal, this again suggests a level dependence in envelope timing transduction cues. This last effect is observed for both high and medium SR fibers.

No ITD-based metrics captured here could provide azimuth cues for the vocoded VAS word stimuli, even though the ILD metric alone could not account entirely for the listener's lateralization performance. This bizarre finding suggests a possible role for spectral shift cues, which were not considered here. Furthermore, no assumptions were made regarding brainstem temporal processing,

which is known to play a role in temporal pattern sharpening, and would likely have affected these outcomes.

The use of BCP patterns themselves as predictors proved an enlightening approach, showing which calculations (inner product, difcor, or sumcor) produced patterns unique to a given azimuth. The results for comparisons of BCP patterns between word and nonword stimuli suggest that broadband stimuli may produce the strongest TFS cues to localization. These illustrations also point out the apparent weakness in the signal representations of envelope ITDs. It seems that the most unique BCPs derive from difcor calculations, and that these calculations, as above, produce no useful information for determining the azimuth of vocoded VAS words. However, the BCP uniqueness for a given azimuth, as encoded in the $\delta_b$ calculation, well explained the difference in lateralization errors between clean and vocoded words. This finding suggests that instead of metric extraction, the binaural auditory system may rely partially upon pattern template matching for lateralization.

Even as much research has focused on the effects of acoustic envelope and TFS manipulations, e.g., speech recognition and vocoding, the connections among acoustic and neural envelope and TFS are not entirely clear. It is apparent that there is neither an exclusive link between acoustic envelope and neural envelope nor between acoustic TFS and neural TFS. We are forced to acknowledge that acoustic decomposition by modulation filtering may be artificial with respect to neural transduction mechanisms and neural representations. As we move forward, it may behoove us to explore a signal decomposition based on neural envelope and TFS definitions. Spectrogram reconstruction from neural signals can be performed via a Lateral Inhibition Network, among other methods. If a neural signal can be decomposed into its envelope and TFS components, then the corresponding envelope and TFS spectrograms could be reconstructed. With separated neural envelope and TFS responses in-hand, one is better prepared to

find the individual contributions of envelope and TFS to perceptual result, such as speech recognition and localization.

Additionally, future studies may benefit from calculations based on short-time windowed neural responses. It is unlikely that neural responses are calculated *en vivo* from the entire historical response of the peripheral auditory system to stimuli of the length used here. It is more likely that the auditory system executes binaural calculations in consecutive, possibly overlapping, time windows. This kind of analysis would produce a time-dependent binaural output and could be used for better understanding the impact of onset and offset features and the contributions of ongoing binaural differences to lateralization percepts.

**Conclusions**

The quest to chart the physiological machinery behind the psychophysical percept of sound source location continues. The physical cues and transduction pathways are quite well understood, and anatomical pathways for binaural processing have been identified. However, the precise neurophysiological mechanisms underlying the extraction of azimuth information from binaural information are yet disputed. The work here with simulated AN data has shown that fairly precise azimuth information provided by physical ILDs may be extracted by the calculation of the difference between the number of spikes on the ANs from the left and right ears. This is a very simple model and is easily implemented through the observed anatomical inhibition pathway. The mechanisms of ITD extraction are far less clear, however, and none of the metrics attempted here (nor combinations thereof) has predicted azimuth with the precision of the behaving human. Especially vexing was attempting to explain the listener's ability to lateralize vocoded VAS word stimuli, which appear to contain no accurate timing cues. Perhaps the central auditory system is able to weigh consistent cues

more heavily for any given listening situation and ignore the inconsistent or confusing cues. Neither explained is the ability of bilateral cochlear implant users to lateralize based on pulse timing ITDs up to several hundred Hz. The assumed absence of cochlear transduction delay with electric stimulation would suggest the inadequacy of the long-time BCP or stereausis model. This study has introduced several computational techniques and metrics which may be useful for these avenues of further investigation towards which the present findings point.

**References**

Bonham, B. H., & Lewis, E. R. (1999). Localization by interaural time difference (ITD): effects of interaural frequency mismatch. The Journal of the Acoustical Society of America, 106(1), 281–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10420622

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., & Grothe, B. (2002). Precise inhibition is essential for microsecond interaural time difference coding. Nature, 417(6888), 543–7. doi:10.1038/417543a

Brughera, A., Dunai, L., & Hartmann, W. M. (2013). Human interaural time difference thresholds for sine tones: The high-frequency limit. The Journal of the Acoustical Society of America, 133(5), 2839–55. doi:10.1121/1.4795778

Goldberg, J., & Brown, B. (1969). Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli : some physiological mechanisms of sound localization . Response of Binaural Tonal Stimuli : to Dichotic Physiological. Journal of Neurophysiology, 32.

Grothe, B., Pecka, M., & Mcalpine, D. (2010). Mechanisms of Sound Localization in Mammals, 983–1012. doi:10.1152/physrev.00026.2009.

Heinz, M. G., & Swaminathan, J. (2009). Quantifying Envelope and Fine-Structure Coding in Auditory Nerve Responses to Chimaeric Speech. JARO - Journal of the Association for Research in Otolaryngology, 10, 407–423. doi:10.1007/s10162-009-0169-8

Jeffress, L. a. (1948). A place theory of sound localization. Journal of comparative and physiological psychology, 41(1), 35–39. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18904764

Joris, P. X. (2003). Interaural time sensitivity dominated by cochlea-induced envelope patterns. The Journal of neuroscience : the official journal of the Society for Neuroscience, 23(15), 6345–6350. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12867519

Louage, D. H. G., van der Heijden, M., & Joris, P. X. (2004). Temporal properties of responses to broadband noise in the auditory nerve. Journal of neurophysiology, 91(5), 2051–65. doi:10.1152/jn.00816.2003

Macpherson, E. a., & Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. The Journal of the Acoustical Society of America, 111(5), 2219. doi:10.1121/1.1471898

Shamma, S. (1985). Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. The Journal of the Acoustical Society of America, 78(November). Retrieved from http://link.aip.org/link/?jasman/78/1622/1

Shamma, S. A. (1989). Stereausis: Binaural processing without neural delays. J Acoust Soc Am, 86(September), 989–1006.

Stern, R. M., & Trahiotis, C. (1995). Models of binaural interaction. In Handbook of Perception and Cognition, Volume 6: Hearing (Vol. 6). New York, NY.

Swaminathan, J., & Heinz, M. G. (2011). Predicted effects of sensorineural hearing loss on across-fiber envelope coding in the auditory nerve a ). J Acoust Soc Am, 129(6), 4001–4013. doi:10.1121/1.3583502

Swaminathan, J., & Heinz, M. G. (2012). Psychophysiological Analyses Demonstrate the Importance of Neural Envelope Coding for Speech Perception in Noise, 32(5), 1747–1756. doi:10.1523/JNEUROSCI.4493-11.2012

Yin, T. C., & Chan, J. C. (1990). Interaural time sensitivity in medial superior olive of cat Interaural Time Sensitivity in Medial Superior Olive of Cat. Journal of Neurophysiology, 64, 465–488.

Zilany, M. S. a, Bruce, I. C., Nelson, P. C., & Carney, L. H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. The Journal of the Acoustical Society of America, 126(5), 2390–412. doi:10.1121/1.3238250

Zilany, M. S. A., & Bruce, I. C. (2006). Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery, (September). doi:10.1121/1.2225512

Zilany, M. S. A., & Bruce, I. C. (2007). Representation of the vowel Õ } Õ in normal and impaired auditory nerve fibers : Model predictions of responses in cats. doi:10.1121/1.2735117

# Chapter 4

# Spatial hearing with speech temporal fine structure cues in bilateral cochlear implant listeners

**Introduction**

Cochlear Implants (CIs) have provided hearing to hundreds of thousands of people worldwide who have severe-to-profound hearing loss. This technology has progressed from the single-electrode implant, which merely provided lip reading cues and sensory indication that sound was present, to present-day multi-channel devices, which can give some users excellent open-set speech recognition in quiet without lip-reading. In order to continue to close the performance gap between CI and normal hearing (NH) listeners, the difficulties that even bilaterally-implanted CI listeners have understanding speech in noise and localizing sounds must be addressed. Hearing with CIs is not equivalent to normal hearing largely because the activation of the auditory nerve by CI electrode currents is not faithful to the

pattern produced by normal physiologic mechanisms (Moore 2003). Notably, electric stimulation lacks the spatial and temporal resolution of NH mechanisms, factors that are known to contribute to the degraded ability of CI users to understand speech in noise and localize sounds (Rubinstein and Miller 1999). Therefore, improving the spatial and temporal resolution of CI electric stimulation should improve users' hearing abilities.

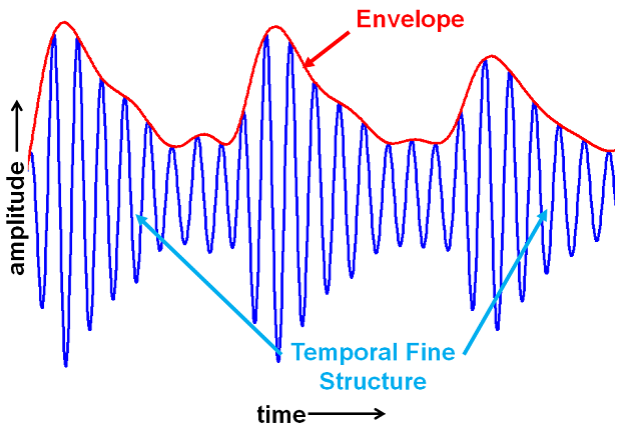



Figure 1. A signal's temporal envelope and temporal fine structure can be calculated for narrowband signals using the Hilbert transform.

A commonly-used signal decomposition technique is to separate narrowband sounds' temporal envelopes from their temporal fine structure (TFS) (Smith et al. 2002). Figure 1 illustrates the distinction between the slowly-varying envelope and its rapidly-varying TFS carrier. It has been found that speech is remarkably robust under signal degradation, and that speech understanding in quiet requires only a few spectrally contiguous channels of envelope information (Shannon et al. 1995). With their relatively small number of independent perceptual channels available (about 7, Friesen et al. 2001), CIs have exploited this fact, and represent sound envelopes as electrode-interleaved, modulated current pulse trains, discarding TFS information (Wilson et al. 1991). Cochlear implants have thus enjoyed much success for the case of speech in quiet, but less so for speech in noise and sound localization.

Much of NH listeners' ability to localize sounds relies heavily on the interaural time difference (ITD) information carried by acoustic TFS at frequencies below 1.5 kHz (Bronkhorst and Plomp 1988; Wightman and Kistler 1992; Macpherson and Middlebrooks 2002; Brughera et al. 2013). Additionally,

the ability of NH listeners to segregate target talkers and competing maskers by location is known to be important for understanding speech in noise (Kidd et al. 1994; Kidd et al. 1998; Hawley et al. 1999; Arbogast et al. 2002; Arbogast et al. 2005; Kidd et al. 2005; Kidd et al. 2005; Ihlefeld and Shinn-Cunningham 2008; Garadat et al. 2009; Kidd et al. 2010; Darwin 2011). This phenomenon, known as spatial release from masking (SRM), also largely depends on receiving ITDs carried by low-frequency TFS, and may rely on the same physiologic machinery as that which suppresses echoes, the precedence effect (Freyman et al. 1999; Hawley et al. 2004; Agrawal et al. 2006; Drennan et al. 2007). In the absence of TFS information with electric hearing, interaural level differences (ILDs) and envelope ITDs are the only binaural cues available to bilateral CI listeners for spatializing sounds, but these cues alone may be inadequate to achieve SRM (van Hoesel and Tyler 2003; Ihlefeld and Litovsky 2012). The discarding of acoustic TFS by most of today's CI processing strategies therefore likely contributes to CI listeners' difficulty in listening conditions other than speech in quiet, and it may be beneficial to reproduce acoustic TFS information in CIs' electrical signals (Wilson et al. 2004).

In NH, TFS is thought to be encoded in the timing of nerve firings that "phase-lock" to mechanical oscillations of the basilar membrane produced by a sound. At electrical stimulation rates below several thousand pulses per second, a CI's current pulses induce "super" phase locking, whereby the auditory nerves fire synchronously with every electric pulse (Kiang et al. 1970; Kiang and Moxon 1972). Therefore, acoustic TFS should be able to be accurately encoded by timing CI stimulating pulses to a given phase of the acoustic signal. Currently in CIs, constant-rate, high-rate (≥ 900 Hz) pulse trains are typically used in order to represent speech envelopes as faithfully as possible (Loizou et al. 2000; Galvin and Fu 2005). While these rates are within the range of phase-locking of the auditory nerve to electric stimulation, they are above the range of pulse timing sensitivity for rate (Carlyon et al. 2010) and ITD discrimination (Laback et al. 2005; Majdak et al. 2006; van Hoesel et al. 2009; Churchill et al. 2012).

Introduction of low-rate (< 300 Hz) “electric TFS” could enable CI listeners access to low-frequency TFS ITD cues, which may be critical for improved sound localization and SRM.

Besides the loss of TFS information, there may be additional negative consequences to using constant-rate pulse trains for electrical stimulation in CIs. Phase-locking of the auditory nerve to constant-rate electric pulse trains removes the natural stochasticity in the timing of nerve firings. This natural randomness is thought to be important for some normal hearing mechanisms, and there have been attempts made to restore it in CIs (Rubinstein et al. 1999; Morse and Meyer 2000; Chatterjee and Robert 2001). Furthermore, the use of constant pulse rates in CIs may lead to adaptation, wherein the nervous system disregards a repetitive, periodic signal, reducing the signal’s information transfer capacity (Smith 1979; Hafter et al. 1983; Laback and Majdak 2008).

Recent advances in processing for CIs have attempted to address these deficits by timing the stimulating electric pulses to represent acoustic TFS information (Zierhofer 2003; Nie et al. 2005; Sit 2007; Sit et al. 2007; van Hoesel 2007). Although listeners often report a preference for strategies which preserve TFS in pulse timing over conventional, constant-rate strategies following their take-home familiarization periods, testing results have not shown significant evidence of any hearing benefits relative to other clinical strategies (Arnoldner et al. 2007; Riss et al. 2008; Riss et al. 2009; Schatzer et al. 2010; Vermeire et al. 2010; Riss et al. 2011). Additionally, although sensitivity to binaural timing cues has been measured in bilaterally-implanted CI users (van Hoesel and Clark 1997; van Hoesel et al. 2002; Laback et al. 2004; Laback et al. 2005; Long et al. 2006; Jones et al. 2008; Laback and Majdak 2008; Smith and Delgutte 2008; van Hoesel et al. 2009), several studies that have investigated psychophysical binaural benefits due to pulse timing derived from speech signals’ acoustic TFS show inconclusive results (van Hoesel and Tyler 2003; van Hoesel et al. 2008). However, these previous investigations into pulse

timing sensitivity may have been confounded by the effects of current spread and the inclusion of ILDs. In the present study, by using direct stimulation techniques, ILD cues were explicitly excluded, leaving pulse timing and envelope ITDs as the only available cues for localization. With the goal of identifying a set of processing parameters which allow for both speech understanding and ITD sensitivity, we systematically examined the effects of channel pulse rate and pulse timing on ITD discrimination, ITD lateralization, and speech recognition of multi-channel speech stimuli. Results from these tests may reveal the necessary parameters for better sound localization and understanding of speech in noise.

Given that ITD sensitivity diminishes at rates above 300 Hz, that ongoing ITDs are most accessible to CI listeners at low rates, and that high-rate pulse trains are superior at faithfully representing speech envelopes, a trade-off appears to exist between speech understanding and ITD sensitivity. The present study has examined this trade-off by testing speech understanding and ITD sensitivity using direct stimulation with eight-channel speech stimuli at three electrical stimulation rate combinations: 1) low rates (100 – 173 Hz) on all electrodes; 2) low rates on four apical electrodes and high rates (894 – 1547 Hz) on four basal electrodes; and 3) high rates on all electrodes. A novel TFS-retaining strategy was tested against a conventional, constant-rate strategy, Continuous Interleaved Sampling (CIS) (Wilson et al. 1991). By testing with both the TFS and CIS strategies, we were able to compare the use of envelope and TFS ITDs across the three rate combinations. In order to test exclusively the effect of pulse timing, pulse rates for a given rate combination have to be the same for both strategies. Thus, for a given rate combination and stimulus token, the TFS and CIS strategies each produced a stimulus with the same average pulse rates.

## Methods

### A. Experimental protocols

Eight bilateral CI users (see table 1) each performed the following three tasks: lateralization of single-word stimuli, left/right discrimination of single-word stimuli, and closed-set speech recognition in quiet using four- or five-word sentences. Four of these listeners (IAJ, IBK, IBR, and ICM) also completed speech-in-noise testing. All listeners were Cochlear device users whose everyday strategy was Advanced Combination Encoder (ACE) and used electrode pulse rates ≥ 900 Hz. Stimuli were presented via direct stimulation with bilaterally-linked research processors (L34s) using the Nucleus Interface Communicator (NIC) software suite by Cochlear. Listener response was collected via touch-screen graphical user interfaces, shown in figure 2, and correct-answer visual feedback was provided following each trial. Two processing strategies and three rate combinations were tested in every task. For the speech recognition task, a given sentence was constructed using one of the six permutations of rate combination and strategy. The presentation order of the six permutation conditions was otherwise randomized among trials for all tasks. A test session for a given task lasted approximately 20 minutes, and the order of tasks conducted was varied so as to support listener alertness and engagement. Informed consent was obtained, and listeners were monetarily compensated for their participation. All procedures were

| Subject | Age | Left Internal/ External | Right Internal/ External | Strategy | Hearing Aid Use | L/R CI Use |
|---|---|---|---|---|---|---|
| IAJ | 68 | CI24M/Freedom | CI24R/Freedom | ACE | 46 y | 17/9 y |
| IBF | 61 | CA/Freedom | CA/Freedom | ACE | 14 y | 5/7 y |
| IBK | 73 | CI24R/Freedom | CA/N5 | ACE | 8 y | 10/4 y |
| IBM | 59 | N5/N5 | CA/N5 | ACE | 16 y | 3/7 y |
| IBN | 66 | CA/Freedom | CI24R/Freedom | ACE | 50 y | 3/13 y |
| IBR | 59 | N5/N5 | CI24R/Freedom | ACE | 22 y | 3/9 y |
| ICD | 56 | CA/Freedom | CI24R/Freedom | ACE | 40 y | 4/10 y |
| ICM | 60 | N5/N5 | CA/N5 | ACE | 29 y | 3/1 y |

Table 1. Data regarding the eight bilateral CI listeners in this study.

approved by the University of Wisconsin's Institutional Review board. All signal processing and data acquisition was executed on custom software with MATLAB.

For the speech recognition task, listeners were asked to identify each of the four or five words in a low-context sentence. Sentences were those spoken by Male #1 from the Kidd et al. (2008) corpus, e.g., "Mike bought five cards," "Jill lost four red hats," etc. Target stimuli were presented from 0° azimuth, i.e., with an ITD of zero. For the lateralization task, listeners identified the perceived azimuthal location of the presented single-word stimulus along a 180° arc. Lateralization stimuli consisted of the names from the above sentence corpus, and applied ITDs were calculated from the head-related transfer functions (HRTFs) of the KEMAR manikin (Algazi et al. 2001) for source azimuths from -70° to +70° in 10° increments for most listeners, and from -50° to +50° in 10° increments and ±90° for several early listeners. For the discrimination task, listeners performed a two-interval, two-alternative forced-choice task, indicating whether the stimulus' perceived location moved from right-to-left or from left-to-right between two presentations of the same speech token, the second interval containing an ITD opposite of that in the first; listeners discriminated positive and negative ITDs of 50, 100, 200, 400, 800, and 1600 μs. Discrimination stimuli also consisted of names from the sentence corpus used in the speech recognition task.

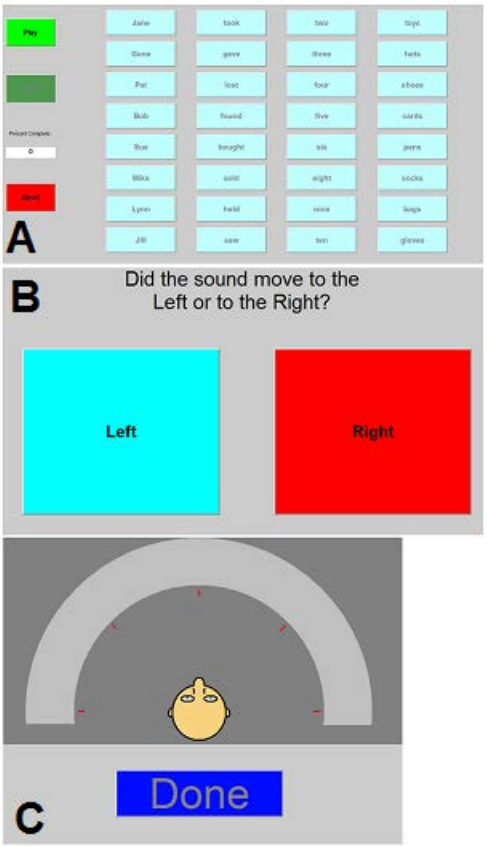



Figure 2. Graphical User Interfaces used for Speech Recognition (A), Discrimination (B), and Lateralization (C) testing. Listeners responded via a touch-screen.

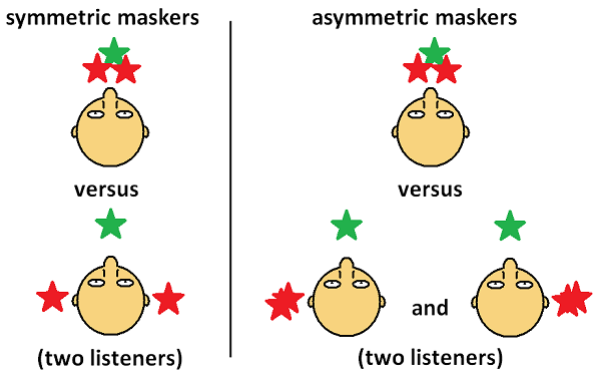



Figure 3. Masking conditions for speech recognition in noise, where ITDs were the only location cues.

Four listeners also completed limited testing of speech in noise in order to assess SRM. As with the speech-in-quiet paradigm above, the target talker was Male #1 from the Kidd et al. (2008) corpus, and was presented with 0 ITD. The maskers consisted of Females #9 and #10 from the same corpus and were presented either collocated with the target (at 0°, or 0 ITD), symmetrically separated from the target (one masker contained the ITD for +90° and the other masker contained the ITD for -90°), or asymmetrically separated from the target (both maskers contained the ITD for either +90° or -90°). These conditions are illustrated in figure 3. Two of the listeners completed the speech in noise testing with symmetrical maskers (IAJ and IBR), and two listeners completed the testing with asymmetrical maskers (IBK and ICM).

Prior to testing, each listener was mapped to find threshold and comfortable current levels for all active electrodes in left and right implants at rates representative of low- and high-rate test stimuli - 150 Hz and 1500 Hz, respectively. This mapping is in many ways similar to the procedure performed by the listeners' audiologists for their clinical device settings. Stimulation consisted of 300-ms trains of monopolar, biphasic current pulses with phase durations of 25μs and an interphase interval of 8 μs. Next, levels were adjusted to produce equal loudness sensations for all comfortable-level stimulation at all rates. This was performed by playing several electrodes sequentially and asking the listener to indicate which intervals needed to be adjusted in order to make them all have the same loudness. This procedure was repeated until all electrodes were played. Finally, levels on each electrode pair were adjusted to produce ILD-centered auditory images. Electrode pairs were activated simultaneously, and

listeners provided feedback to the experimenter as to whether the auditory image was centered or to the left or right, following which the experimenter would adjust the levels to produce a more centered auditory image. Because place-pitch matching of left/right electrodes has been found to be important for ITD sensitivity (Kan et al. in press; van Hoesel et al. 2008; Poon et al. 2009; Litovsky et al. 2010), eight binaurally pitch-matched electrode pairs were selected for the presentation of speech stimuli. Pitch-matched electrode pairs were selected based on data collected from these listeners in previous experiments, wherein the listeners performed direct pitch comparisons for many pairs of electrodes.

For speech recognition in quiet, 100 word trials (20 or 25 sentence trials) were collected for each strategy and rate combination. Resulting percent correct scores were arcsine-transformed (Studebaker, 1985) prior to analyses. For ITD discrimination, at least 40 repetitions were collected for each strategy and rate combination at each of at least 4 ITDs. Logistic function psychometric curves (Wichman and Hill, 2001) were fitted through ITD discrimination percent correct points to calculate just-noticeable differences (JNDs, 71% correct) for each listener and also for pooled response data. Some listeners showed no ITD discrimination sensitivity with the CIS strategy for several rate combinations, and no JNDs were calculable. For ITD lateralization, 10 repetitions were collected for each strategy, rate combination, and azimuth. Linear best-fit psychometric functions (response azimuths as functions of input azimuths) were calculated for each listener and also for pooled responses. The slopes of these input-output functions characterize the listeners' abilities to use the available cues and are used here as the primary metrics of lateralization ability. Speech-in-noise data were collected for each masker position, rate combination, strategy, and for at least 4 signal-to-noise ratios (SNRs) for each listener. Listeners IAJ, IBK, IBR, and ICM respectively completed 48, 40, 36, and 16 repetitions for each combination of conditions. Speech reception thresholds (SRTs) were calculated from best-fit logistic

function psychometric curves at 50% correct speech recognition. Spatial release from masking was calculated as the difference in SRTs between separated and collocated masking conditions.

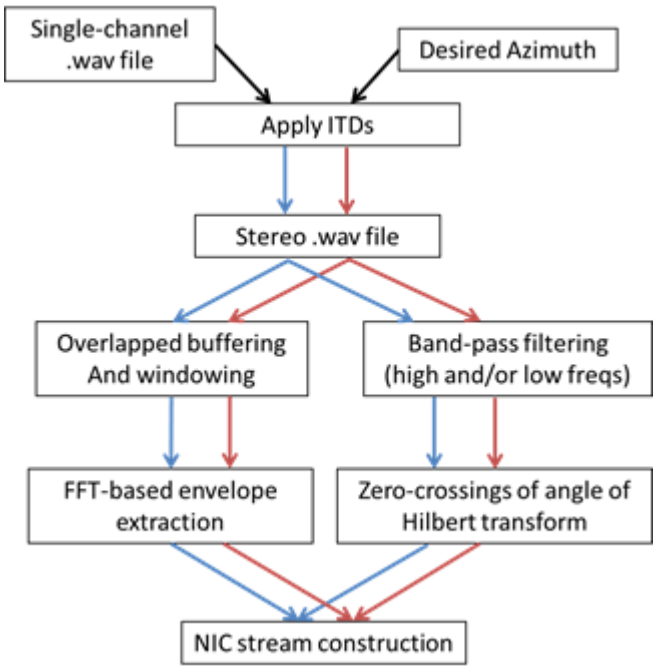



Figure 4. Block diagram for the TFS processing strategy. Rates used for the CIS strategy's isochronous stimulating pulse trains were the average rates of the pulse trains obtained for the corresponding speech token processed by the TFS strategy.

**B. Processing strategies**

Figure 4 shows a block diagram representing the signal processing steps described below. Following stereophonication and the application of ITDs between left and right channels, stimuli were resampled from 44 kHz to 100 kHz, and henceforth each stereo channel was processed separately. This independence of left/right processing would allow for implementation on non-linked processors, but the assurance of controlled, synchronous stimulation would be lost. As described below, envelopes and TFS were extracted using separate techniques.

A short-time Fourier transform-based method was used for envelope extraction. First, stimuli were buffered into 512-point (5.12 ms) slices with 256-point (2.56 ms) overlaps between adjacent time slices. Next, a 512-point Blackman window was applied on each time slice. The Fourier transform was then performed on each time slice, and the frequency-domain signal between bins 1 and 82 was divided into 9 logarithmically-spaced corner frequencies, resulting in 8 channels approximately logarithmically spaced between 200 Hz and 16 kHz. Table 2 shows these envelope channel corner frequencies. Signal

| Frequency (Hz) |
|---|
| 195 |
| 390 |
| 781 |
| 1171 |
| 1953 |
| 3125 |
| 5468 |
| 9375 |
| 16015 |

Table 2. Nine logarithmically-spaced corner frequencies were calculated for extraction of eight channels' envelopes based on 512-point overlapping windows and 100 kHz sampling frequency.

magnitudes were calculated for each channel and time slice, and then up-sampled by a factor of 256, resulting in a 100 kHz-sampled envelope for each of 8 channels and for each ear.

Pulse timing information was extracted from the TFS of band-pass-filtered signals for each ear. First, signals from each side were band-pass filtered into two channels- a low-rate and a high-rate channel. The low-rate channel had corners at 100 Hz and 173 Hz, and the high-rate channel had corners at 894 Hz and 1547 Hz. These corners were based on a logarithmic spacing of 8 channels between 100 Hz and 8 kHz, and were not directly related to the spacing of channels for envelope extraction. Filtering was performed with third-order Butterworth filters using forward-and-reverse filtering, a zero phase-shift technique which doubles the effective filter order. Next, a Hilbert transform was performed on the output of each of the two channels on each side:

$$H[x(t)] = \frac{1}{\pi}\int_{-\infty}^{\infty} dt' \frac{x(t')}{t - t'}$$

Finally, the positive-going zero-crossings of the Hilbert-transformed signal were extracted for each channel, giving 100-kHz sampled vectors consisting of ones at the channel's Hilbert transform positive-going zero-crossings, and zeros elsewhere. The positive-going zero-crossings of the Hilbert transform correspond to the peaks in the original band-pass filtered signal and therefore represent a good measure of the TFS present in the respective filters' output signals. The constant-rate pulse timing vectors for the corresponding CIS stimuli were created by dividing the sum of the TFS-based pulse timing vectors by their length, giving average pulse rates, from which constant-rate pulse timing vectors of the same length and average pulse rate were created.

Pulse timing vectors (high and/or low rate, TFS-based or constant-rate CIS) were then modulated by the appropriate envelopes based on the strategy and rate combination to be presented on a given trial. For the low-rate stimuli, all eight stereo channels' envelopes modulated the corresponding left and right low-rate pulse timing vectors. For the high-rate stimuli, all eight channels' envelopes modulated the corresponding left and right high-rate pulse timing vectors. For the mixed-rate stimuli, the four apical (low frequency) channels' envelopes modulated the low-rate pulse timing vectors, and the four basal (high frequency) channels' envelopes modulated the high-rate pulse timing vectors. Modulated pulse timing vectors were then resampled into 70-μs wide bins, uniformly interleaved, and power-law compressed (exponent = 1/3) into current levels between the listeners' threshold and comfort levels. Figure 5 shows an example electrodogram segment from a mixed-rate, TFS-strategy stimulus token for one ear. Envelopes of speech signals are therefore presented bilaterally on eight pitch-matched electrode pairs, with pulse timing for the TFS strategy containing relevant ITD information, and pulse timing for the CIS strategy containing a constant ITD of 0 at a constant rate. Stimuli were generated prior to listeners beginning the experiments.

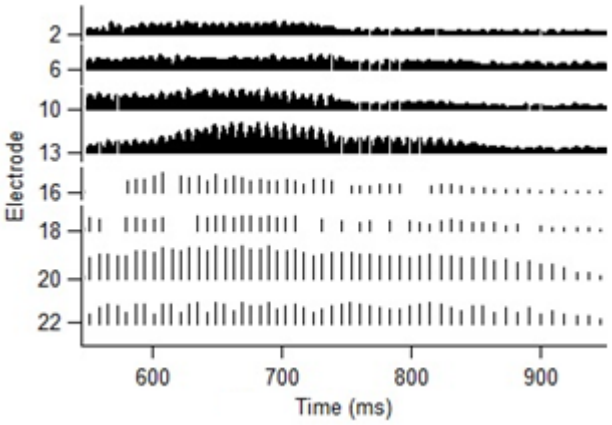



Figure 5. Electrodogram showing pulse timing and amplitudes for a segment of a mixed-rate, TFS-strategy stimulus token.

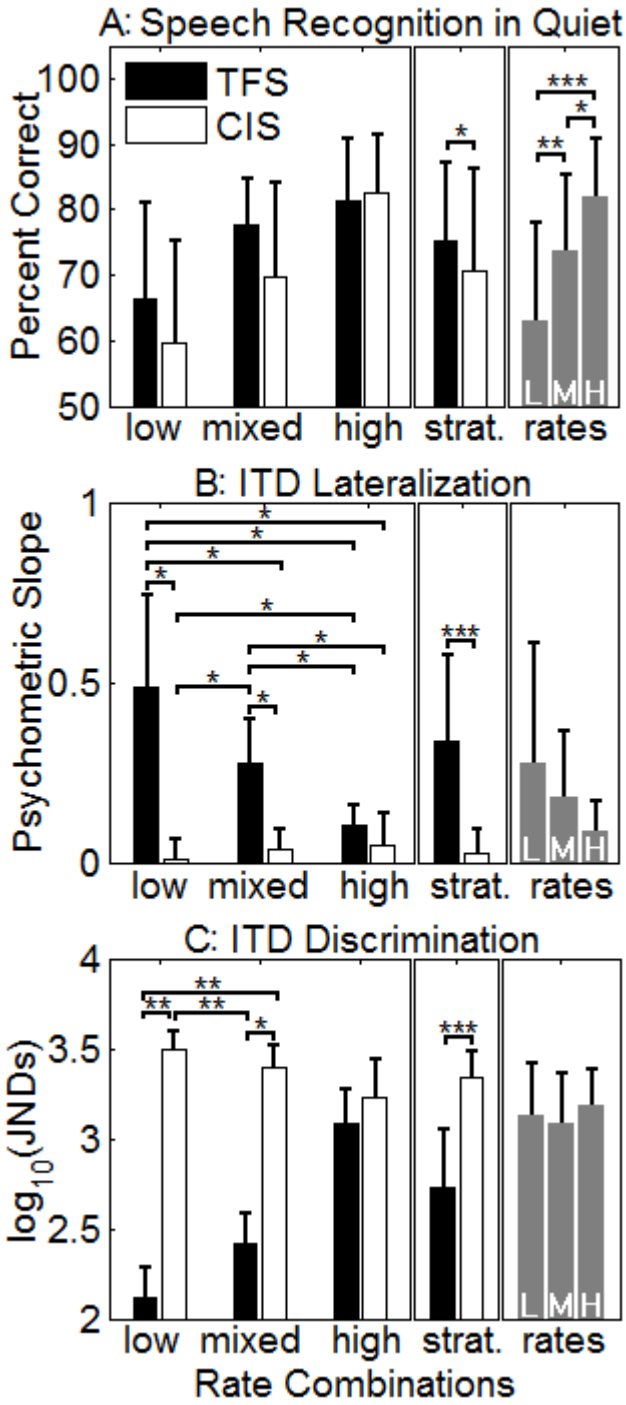



Figure 6. A) Percents correct speech recognition in quiet; B) ITD lateralization psychometric function slopes; and C) ITD discrimination JNDs shown for each strategy/rate combination, averaged across rates, and averaged across strategies. Error bars show the standard deviation of measured listener scores. Asterisks indicate significant differences:
* - $P < 0.05$; ** - $P < 0.01$; *** - $P < 0.001$

**Results**

Data were collected and analyzed as described above. Figure 6 shows results from all three tasks in quiet. Listeners' metric averages are shown for each strategy and rate combination, for each strategy overall, i.e., averaged across rate combinations, and for each rate overall, i.e., averaged across strategies. Percent correct scores for speech recognition are plotted in panel A of figure 6. Figure 7 shows the percent correct scores for each listener, rate combination, and strategy, and figure 8 shows the confusion matrices for speech in quiet. Separate confusion matrices for each strategy and rate combination are similar to each other and hold no novel revelations. Listeners tended to have higher scores with the TFS strategy for low and mixed-rate stimuli, whereas scores tended to be roughly equal for high-rate stimuli. As seen also in figure 7, the variability in listener performance was higher for low and mixed rates than for high rates, with

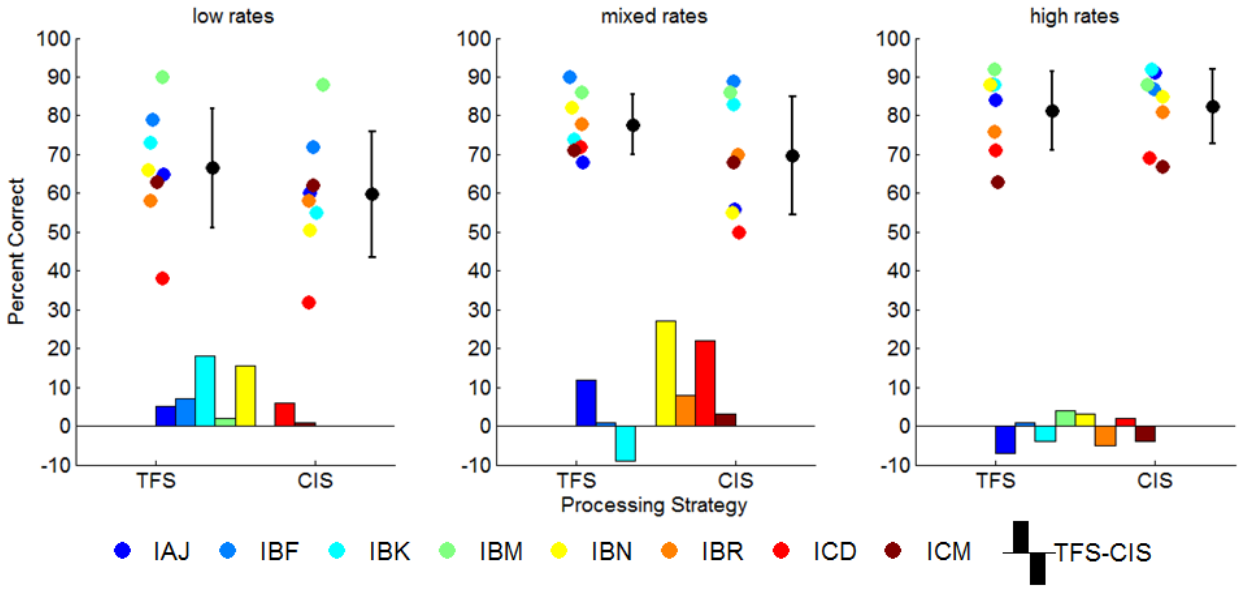



Figure 7. Percent correct speech recognition in quiet. Colored markers indicate individual listener scores. Black markers indicate average scores for a given rate combination and strategy, and error bars show standard deviation of listener scores. Bars at the bottom of each plot show the difference between TFS and CIS scores for each listener and rate combination.

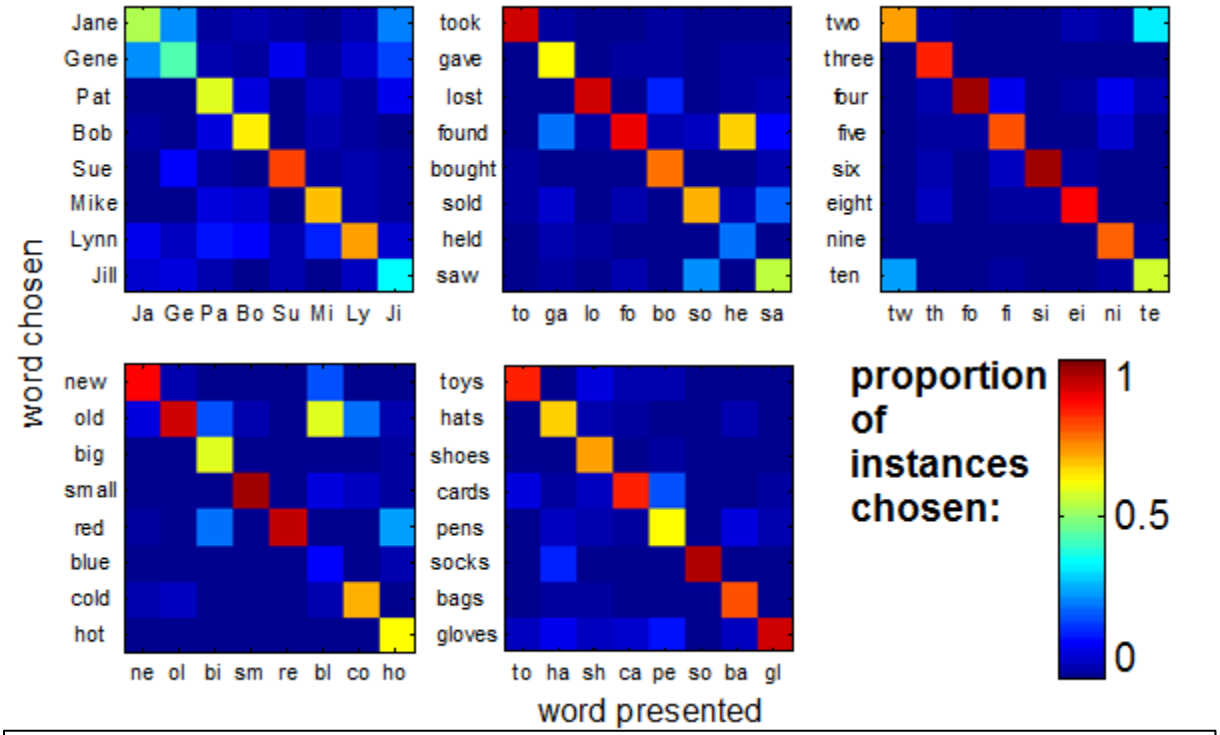



Figure 8. Confusion matrices for speech in quiet. The proportion of times that a listener selected a specified word when a given word was presented is shown by a color scale. The strong diagonal indicates that listeners often chose the correct words.

some listeners' performance at low or mixed rates nearly equal that for high rates and some listeners' performance reduced at low or mixed rates. A planned, two-tailed, paired t-test of scores grouped by strategy indicated that scores with the TFS strategy were higher overall than those with the CIS strategy ($p$ = 0.029). However, this was largely driven by differential performance between the strategies at low and mixed rates. Additional two-tailed, paired t-tests of scores grouped by strategy indicated that performance with the TFS strategy was better when considering only low and mixed rates ($P$ = 0.007) or only low rates ($P$ = 0.018). Planned, two-tailed, paired t-tests of scores grouped by rate combination indicated that scores with high-rate stimuli were higher than with mixed ($p$ = 0.012) or low rates ($p$ < 0.001) and that scores with mixed stimulation rates were higher than scores with low rates ($p$ = 0.002). A two-way, within-subjects analysis of variance (ANOVA) was conducted on arcsine-transformed speech recognition scores for strategy and rate combination. This analysis revealed a significant main effect of rate combination only [$F_{(2,14)}$ = 12.8, $P$ < 0.001]. Post-hoc two-tailed, Bonferroni-corrected, paired t-tests, conducted between all possible pairs of strategy/rate combination, revealed only that low-rate CIS speech was less intelligible than high-rate CIS-processed or TFS-processed speech ($P$ < 0.05).

Average lateralization psychometric function slopes are shown in panel B of figure 6. Individual listeners' average responses at each azimuth are shown in figure 9, along with psychometric function fits through pooled responses. Positive and significant slopes ($P$ < 0.01) were observed for all listeners with low- and mixed-rate TFS-processed stimuli, indicating that listeners were able to use low-rate pulse timing for lateralization. The response bias evident for some listeners and rate combinations in figure 9 is most likely a result of residual ILDs following the qualitative loudness balancing. A planned, two-tailed, paired t-test of lateralization slopes grouped by strategy indicated that slopes with the TFS strategy were higher overall than those with the CIS strategy ($p$ < 0.001). Planned t-tests of slopes grouped by rate combination revealed no differences. A two-way, within-subjects ANOVA was conducted on

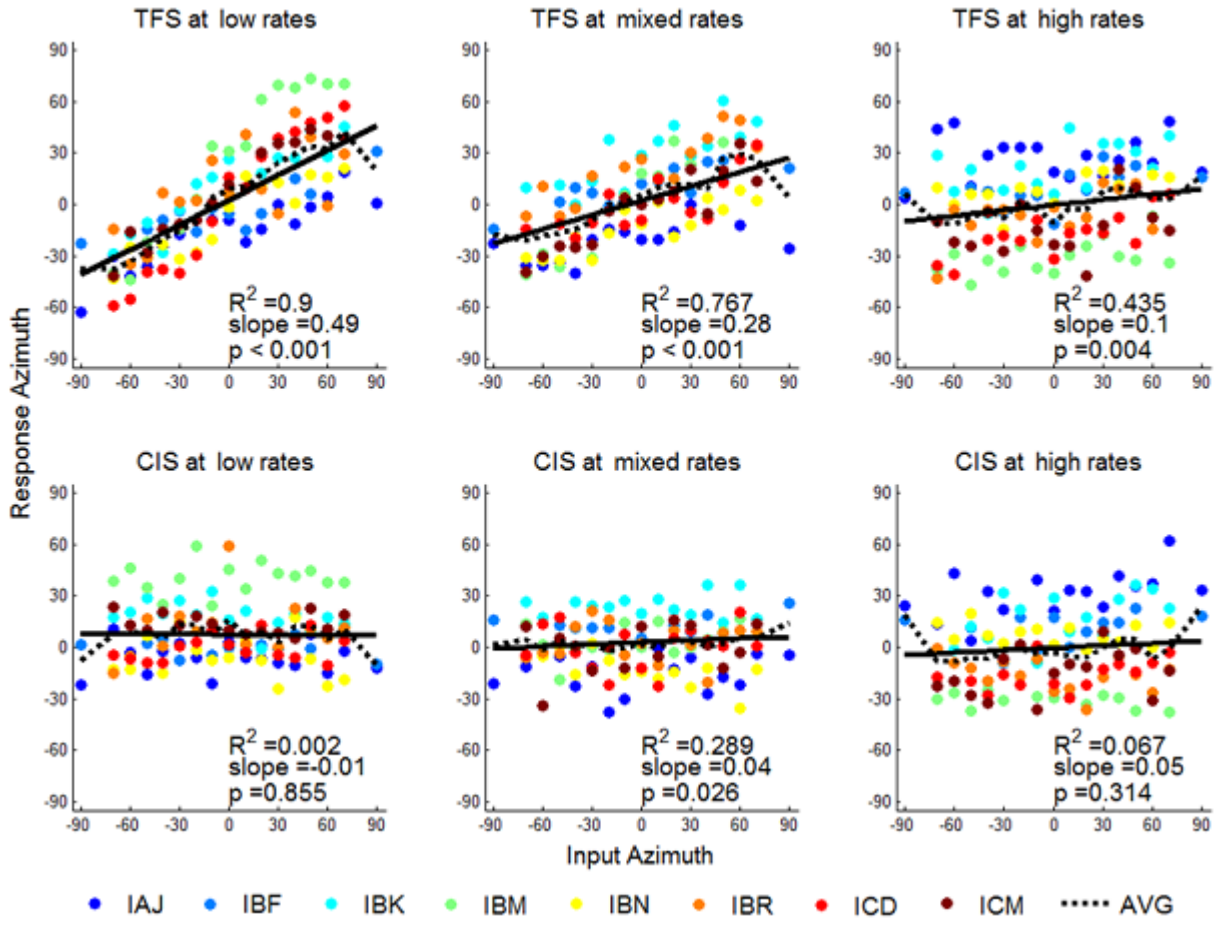



Figure 9. Average lateralization responses for each listener, strategy, rate combination, and presented azimuth. The dashed line shows the average of all listener responses and the solid line shows the best-fit linear psychometric curve though pooled listener data, with accompanying statistics.

lateralization slopes for strategy and rate combination. This analysis revealed significant main effects of strategy [$F_{(1,7)} = 47.1$, $p < 0.001$] and rate combination [$F_{(2,14)} = 10.0$, $p = 0.002$], and a significant interaction of strategy and rate combination [$F_{(2,14)} = 26.2$, $p < 0.001$]. Post-hoc two-tailed, Bonferroni-corrected, paired t-tests between all possible combinations of strategy and rates (15 total) revealed that low-rate TFS-processed stimuli resulted in higher lateralization slopes than any condition besides mixed-rate TFS ($p < 0.05$), that mixed-rate TFS-processed stimuli resulted in higher lateralization slopes than

any condition besides low-rate TFS ($p < 0.05$), and that high-rate TFS-processed stimuli resulted in higher lateralization slopes low-rate CIS-processed stimuli ($p < 0.05$).

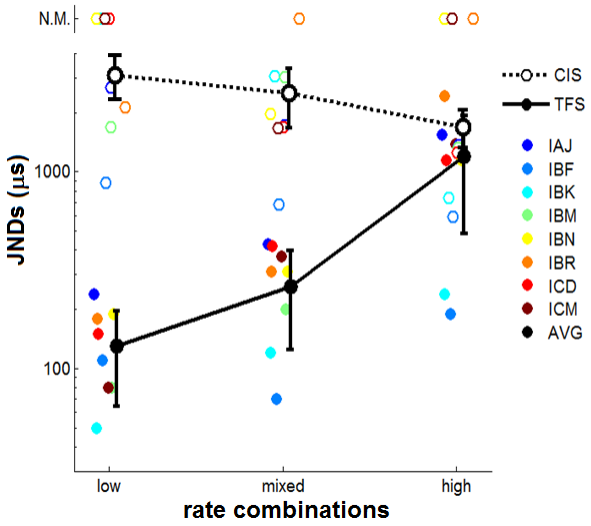



Figure 10. Just-noticeable differences in ITD discrimination. Calculated JNDs from pooled listener response data (in black) are connected by solid and dotted lines for TFS and CIS strategies, respectively. Since listeners discriminated positive from negative ITDs, the magnitude of differences detected is double the values shown here.

Panel C of figure 6 shows average JNDs. Individual listeners' JNDs are plotted in figure 10. Since listeners discriminated opposite-signed ITD pairs, the actual magnitudes discriminated were double those shown. At low and mixed rates, listeners had significantly lower JNDs with TFS-processed stimuli than with CIS-processed stimuli, but these JNDs converged at high rates. A planned, two-tailed, paired t-test of JNDs grouped by strategy indicated that JNDs with the TFS strategy were lower overall than those with the CIS strategy ($p < 0.001$). T-tests of JNDs grouped by rate combination revealed no differences. A two-way, within-subjects ANOVA was conducted on discrimination JNDs for strategy and rate combination. This analysis revealed a significant main effect of strategy [$F_{(1,7)} = 66.3$, $p < 0.001$] and a significant interaction of strategy and rate combination [$F_{(2,14)} = 9.2$, $p = 0.003$]. Post-hoc two-tailed, Bonferroni-corrected, paired t-tests between all combinations of strategy and rates revealed that low-rate, TFS-processed stimuli produced lower discrimination JNDs than either low-rate or mixed-rate CIS-processed stimuli ($p < 0.01$), and that and mixed-rate, TFS-processed stimuli produced lower JNDs than either low-rate ($p < 0.01$) or mixed-rate ($p < 0.05$) CIS-processed stimuli.

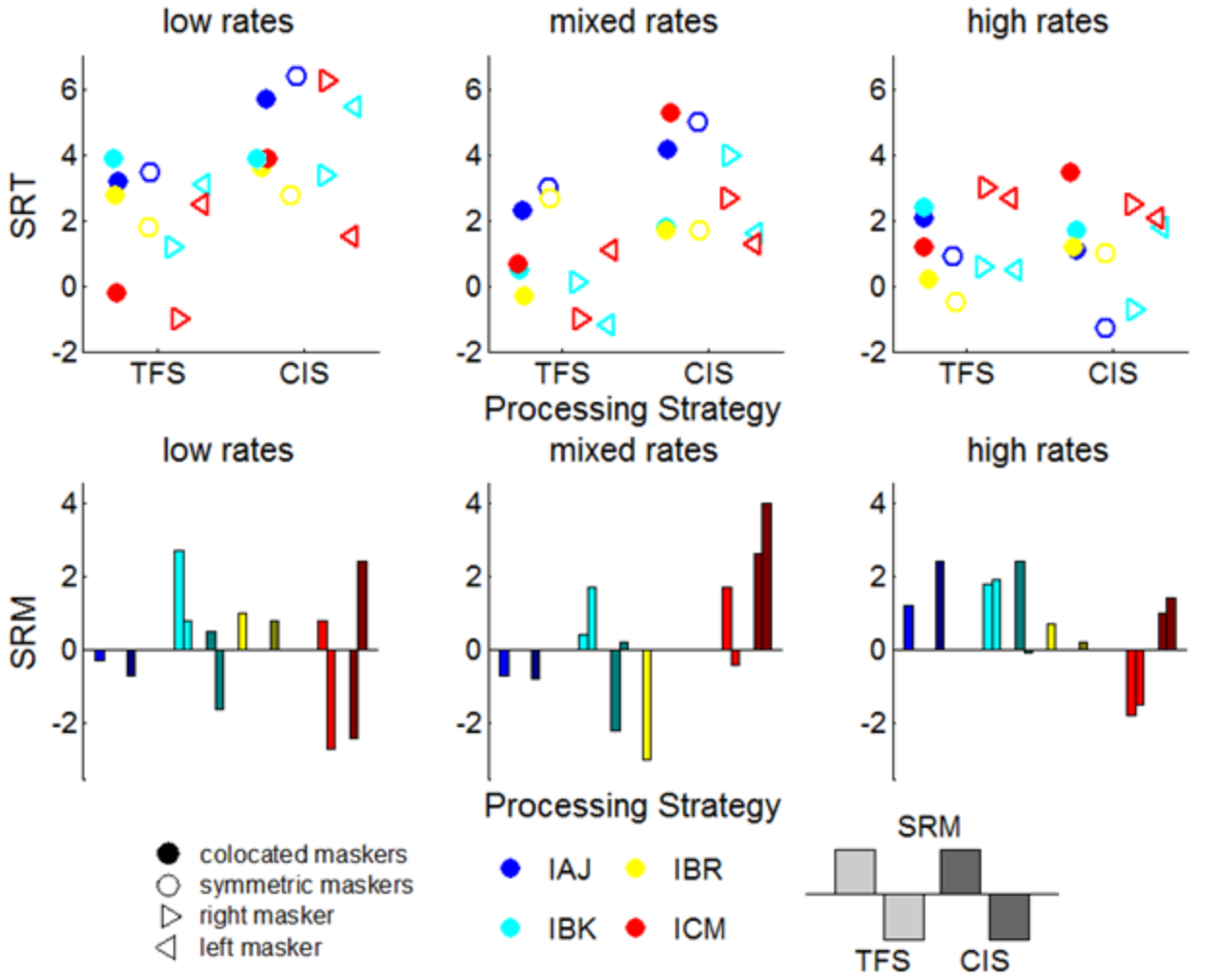



Figure 11. Speech reception thresholds and spatial release from masking (in dB SNR) are shown for each listener, strategy, and rate combination. For each listener as applicable, SRM is shown for TFS (light-colored bars) and CIS (dark-colored bars) strategies in the following order from left to right: symmetrically-separated maskers, masker to the right, masker to the left.

Speech reception thresholds from speech-in-noise data are shown in the top three plots of figure 11. Spatial release from masking, calculated as the difference in SRTs between separated and collocated masking conditions, is shown in the bottom three plots of Figure 11. Mirroring the speech-in-quiet percents correct, SRTs were lower (better) overall at high rates, and lower for the TFS strategy than for the CIS strategy at low and mixed rates. An ANOVA was conducted on speech-in-noise SRTs for strategy and rate combination. This analysis revealed significant main effects of strategy [$F_{(1,54)}$ = 14.2, *P*

< 0.001] and rate combination [$F_{(2,54)}$ = 8.5, *P* < 0.001] and a significant interaction of strategy and rate combination [$F_{(2,54)}$ = 3.7, *P* < 0.05]. Bonferroni-corrected pairwise t-tests revealed that SRTs were lower with the TFS strategy than with the CIS strategy at low rates (*P* = 0.023) and mixed rates (*P* = 0.033) and that SRTs were overall higher at low rates than with mixed (*P* = 0.007) or high rates (*P* = 0.009). A two-tailed, paired t-test revealed that SRTs were lower overall with the TFS strategy (*P* < 0.001). Although the average level of SRM was positive, it was not significantly greater than zero. There were no observed effects of masker location on SRT nor of any variable on SRM.

The listeners had no familiarization exposure to the stimuli prior to testing. In order to assess learning effects on the speech recognition task in quiet, the response data for each listener were organized chronologically and a cumulative correct function (CCF) was generated as a function of trial number t for each strategy/rate combination condition and for each strategy summed across rate combinations:

$$CCF(t) = \sum_{t'=1}^{t} TC(t')$$

where *TC(t')* (trial correct) is equal to one if the listener responded correctly on trial *t'* and zero otherwise. The following three models were used to fit the CCFs:

$$\boldsymbol{A})\ CCF(t) = \beta_l \times t$$

$$\boldsymbol{B})\ CCF(t) = \beta_q \times t^2$$

$$\boldsymbol{C})\ CCF(t) = \beta_l \times t + \beta_q \times t^2$$

where $\beta_l$ and $\beta_q$ are the linear and quadratic coefficients, respectively. Model A represents the CCF as a pure linear function in trial number t, and ignores any improvements in performance over time. Model B

represents the CCF as a pure quadratic function in trial number and assumes no base level of understanding. Model C represents the CCF as a sum of linear and quadratic terms in trial number. It assumes a base of speech understanding would provide a linear increase in number correct with trial number and learning effects are represented by a quadratic term in trial number. Two-tailed t-tests of model coefficients found that both models A and C showed higher $\beta_1$ values for the TFS strategy in the mixed-rate condition ($P$ = 0.0486 and $P$ = 0.0036, respectively). These results suggest that there was no significant difference in rates of learning the two strategies, but that speech understanding was easier with the TFS strategy than the CIS strategy in the mixed-rate condition, a prediction reflected in the final percents correct for these conditions.

Listeners performed equally or had better performance with TFS strategy stimuli versus CIS strategy stimuli on all tasks and in all conditions. The TFS strategy resulted in lower discrimination JNDs, higher lateralization function slopes, higher percents correct for speech understanding in quiet, and lower SRTs for speech in noise. Figure 12 shows the various pairwise comparisons of scores for each of the tasks in quiet. Although it is unclear whether ITD discrimination and lateralization rely upon the same neural mechanisms, panel A of figure 12 illustrates a strong relationship between lateralization and discrimination abilities; higher slopes correlate with lower JNDs ($R^2$ = 0.454). Best overall spatial hearing ability is indicated by data points in the lower-right corner of panel A. Panels B and C illustrate the tradeoff between better ITD sensitivity at low rates and superior speech recognition at high rates. In panel B, the ideal combination of scores, low JNDs and high speech recognition percent correct, is again indicated by points in the lower-right corner of the plot. In panel C, higher lateralization slopes and speech recognition scores are in the upper-right corner. From these graphs, it is evident that the TFS strategy more closely approaches these “ideal” performance regions than the CIS strategy, and that for

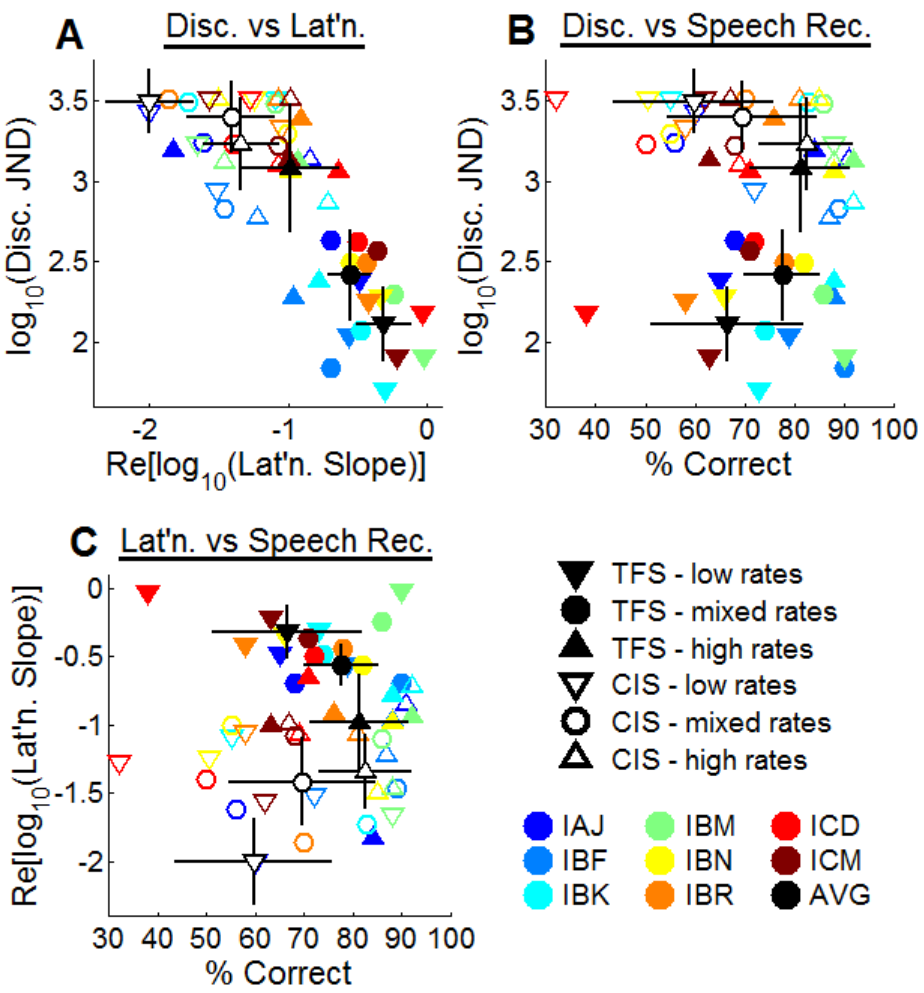



Figure 12. Relationships among measures from all three tests. Colored markers indicate individual listeners' scores, while black markers with error bars indicate average scores from pooled listener response data. Error bars show the standard deviation of measured listener scores. Panel A shows ITD discrimination JNDs versus ITD lateralization psychometric function slopes, panel B shows discrimination JNDs versus percent correct speech recognition in quiet, and panel C shows lateralization slopes versus speech recognition scores.

several listeners, a low- or mixed-rate TFS strategy may provide the optimal cue set for spatial hearing and speech recognition in quiet.

The observed trend in SRTs strongly reflects the trend for percent correct speech recognition in quiet. Namely, listeners generally perform better at high rates and with the TFS strategy when low or mixed rates are used. However, the contribution of improvements in spatial hearing to improvements in speech understanding in noise is not revealed here. Only one listener (IBK) showed consistently positive SRM for the TFS strategy at all rate combinations.

**Discussion**

The results reported here show that redundant, low-rate pulse timing on multiple channels can carry useful ITD information for lateralization and discrimination of speech stimuli with bilateral CIs, even when mixed with high rates on some channels (see figure 6 and 12). The results also show that this useful pulse timing information can be calculated directly from signals' TFS. Furthermore, the results characterize the trade-off between spatial hearing abilities and speech recognition as different pulse rates and processing strategies are used. From the depiction of this trade-off in panels B and C of figure 12, it is clear that the TFS stimuli are superior to CIS stimuli for providing listeners with both speech and spatial cues. However, a single strategy and rate combination did not stand out as the universal best parameter set. One listener's best strategy/rate combination appears to be low-rate TFS, while another's is mixed-rate TFS. Ideal parameter settings are not consistent across all CI users and listening conditions, but these results suggest that bilateral CI listeners may benefit from the inclusion of low-rate or multi-rate, TFS-timed strategies with their clinical maps.

While it is apparent that these listeners could lateralize stimuli due to their sensitivity to ITDs in the pulse timing, it is important to note that we have deliberately ignored the dominant spatial cue for CI listeners, ILDs. Three listeners (IAJ, IBF, and IBK) also completed lateralization testing in which the stimuli contained the full set of lateral cues due to the KEMAR HRTFs, i.e., ITDs and ILDs. Average listener responses for these and ITD-only lateralization tests are shown in figure 13, and average linear psychometric function slopes from this small data set are shown in figure 14. The strategies used for this full-HRTF-cues subset of testing were different from those tested in the other tasks; this CIS strategy used 1000 Hz constant-rate pulse trains on all eight channels, and this TFS strategy used pulse timings calculated independently for each channel from the TFS in that channel. Since this TFS pulse timing

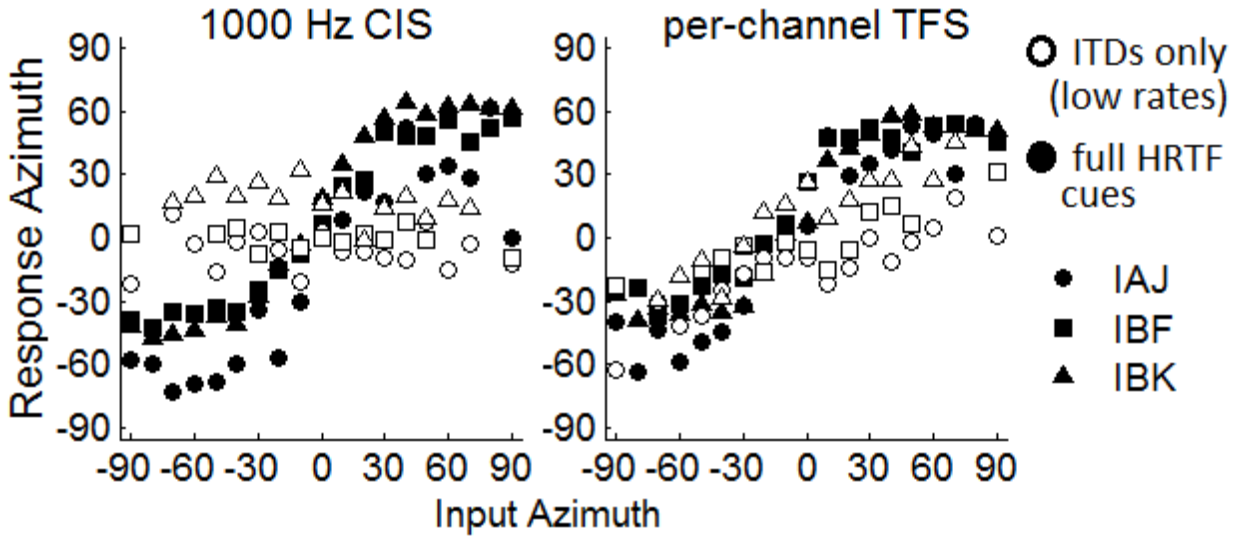



Figure 13. Average lateralization responses from three listeners who completed lateralization of stimuli with ITD cues only and also with all HRTF cues. Open markers indicate responses to stimuli containing only low-rate ITD cues (from testing reported above), and filled markers indicate responses to stimuli containing full HRTF cues. The CIS strategy with full HRTF cues used 1000 Hz pulse trains on all electrodes, and the TFS strategy with full HRTF cues used pulse timing derived uniquely from the TFS of each channel.

calculation produced higher-rate and unrelated pulse trains, pulse timing cues may not have been as accessible with this TFS strategy. Listener response patterns for the two strategies tested here are nearly indistinguishable, indicating that pulse timing ITDs did not contribute to lateralization ability.

Lateralization responses to stimuli with ITD and ILD cues result in steeper psychometric function slopes than those due to stimuli that only contained ITD cues, but they also exhibit the characteristic hemispherical sensitivity displayed by CI listeners in free-field studies, a pattern possibly to be broken by the inclusion of

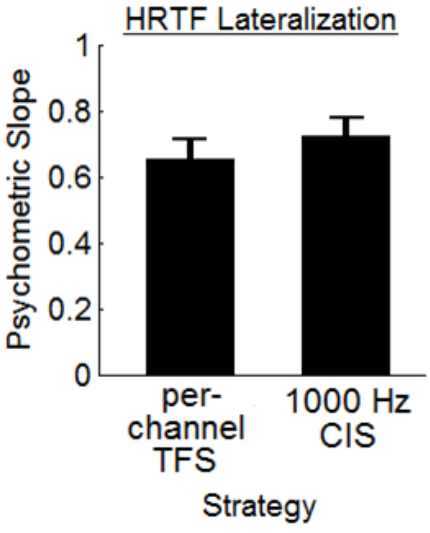



Figure 14. Average lateralization psychometric function slopes for stimuli with full HRTF cues, i.e., ITDs and ILDs, obtained from three listeners who also completed the above ITD-only lateralization experiments. This CIS strategy used 1000 Hz pulse trains on all electrodes, and the TFS strategy used pulse timing derived uniquely for each channel.

prominent and redundant low-rate ITD cues in the pulse timing.

Free-field and direct-connect studies have previously shown that for speech in noise, bilateral CI listeners benefit from having two implants in several ways (van Hoesel and Tyler 2003; Schleich et al. 2004; Litovsky et al. 2006; Litovsky et al. 2009; Loizou et al. 2009). First, they may attend to their better ear, the ear with a more favorable SNR, taking advantage of monaural head shadow effects for 4-5 dB of improvement. Second, some listeners may also exhibit an additional 1-2 dB binaural benefit. Since speech TFS cues are normally unavailable, this binaural benefit is thought to derive from ILDs and/or envelope ITDs. Whereas NH listeners show robust SRM with symmetrically-separated maskers, other work has shown that bilateral CI users perform poorly on conditions in which maskers are symmetrically distributed and monaural head shadow cues are minimal or absent (Agrawal 2008; Misurelli and Litovsky 2012). Some studies have also compared listeners' ILD and ITD sensitivities, finding that bilateral CI listeners are vastly more sensitive to natural ILD cues than natural ITD cues (van Hoesel and Tyler 2003; Litovsky et al. 2010). Direct-stimulation experiments by Long et al. (2006) and Lu et al. (2010) found that binaural unmasking of non-speech signals could be achieved with envelope decorrelation alone, but did not investigate the effects of TFS. Loizou et al. (2009) examined SRM in bilateral CI listeners by presenting speech and informational masker stimuli through bilaterally-linked research processors, but found no binaural advantage for SRM with a conventional constant-rate strategy. The authors suggested that this lack of binaural advantage arises due to poor ITD sensitivity, poor spectral resolution, and/or binaural mismatch. The current study has attempted to compensate for binaural mismatch, and the enhancement of ITD sensitivity with our TFS stimuli may provide the necessary cues for SRM. Van Hoesel, et al. (2008) investigated the use of target ITDs for speech unmasking with several processing strategies, including one that represented TFS in pulse timing at several low-frequency apical electrodes. This strategy, peak-derived timing (PDT), preserves TFS cues by timing pulses to the positive

peaks in the output of each channel's filter (van Hoesel 2007). The study found no binaural speech unmasking when applying a 700-μs ITD to the target signal in the presence of a masker presented from the front (ITD of 0 μs). That study also investigated free-field lateralization of click train stimuli, and found no significant differences in performance among PDT, ACE, and CIS strategies. However, the inclusion of ILD cues in free-field lateralization testing likely masked the impact of pulse timing ITDs. Thus, ILD cues may be more usable for binaural unmasking because CI users appear to be more sensitive to ILDs when either ILD or ITD cues are isolated and presented with fidelity, and because ILD cues are currently more readily available in clinical speech processors. As mentioned in the conclusion of van Hoesel, et al. (2008), several details of the implementation of the PDT strategy may have adversely affected the utility of pulse timing for ITD sensitivity: 1) the use of 19 channels virtually ensures that current spread will produce channel cross-talk, blurring the presentation of ITDs from adjacent electrodes; 2) the use of channel-unique derivation of pulse timing and placing the lowest filter corner frequency at 250 Hz allows only a few of the most apical electrodes to carry ITDs at usable low rates; and 3) even usable (low-rate) ITD information is presented by unique pulse trains at each electrode, and therefore cues may be inconsistent, a serious confound if current spread is considered, in which case adjacent channels' pulse timing may have provided conflicting or confusing cues to surviving auditory nerve populations in a region of current spread overlap. Hence, ITD information presented to listeners in that study may not have been carried under optimal conditions, and superior conveyance of ITDs may have provided more usable cues. Despite the apparent dominance of ILDs in CI spatial hearing, the addition of accessible and possibly redundant ITD information may provide the necessary cues for stream segregation and SRM (Ihlefeld and Litovsky 2012).

The current results have shown stronger evidence for the binaural benefit of TFS-timed pulses than these earlier experiments. This may be due to differences in methodology or signal processing.

First, the deleterious effects of spectral mismatch across ears due to electrode placement and current spread are thought to have been major factors in the observed reduced ability of bilateral CI listeners to utilize deliberately-presented interaural TFS information (Kan et al. in press; Poon et al. 2009). The present study attempted to minimize these effects by stimulating on pitch-matched electrode pairs. Furthermore, the use of only eight electrode pairs allowed for the physical separation of active electrodes, in order to reduce the possible effects of channel interaction due to current spread. Additionally, the present study avoided the potential confounds associated with the presentation of usable ITDs mentioned above by using low-rate (< 200 Hz) pulse timing cues which were redundant across multiple electrodes. The data presented in figure 13 and discussed above were collected using a TFS strategy that was very similar to PDT in that each electrode's pulse timing was determined from the acoustic TFS in the corresponding channel. The similarity of results with these two strategies suggests that in each case, CI listeners were not benefiting from the added pulse timing information, the most likely impediments to using pulse timing ITDs being those noted above, namely, the presence of high rates in most of the channels and that the two channels carrying usable (<500 Hz) pulse timing information, the two most apical channels, were carrying the ITD cue at different rates. Overlap between these two channels due to current spread could easily have led to false or confusing cues.

The use of direct stimulation allows for excellent control of variables, and fills the gap between experiments using direct stimulation with modulated signals and free field experiments with clinical processors on a spectrum of realism. Among binaural CI direct stimulation experiments, the present one falls on the "realistic" side of that spectrum in the following ways. First, the CI signal processing is carried-out independently for left and right channels, and could theoretically be implemented on unlinked processors. Second, the use of speech stimuli for discrimination and lateralization reflects that speech is among the most commonly encountered acoustic signals of interest outside of the testing

booth. Though its understanding can be achieved even when highly degraded, speech is a complex sound, with myriad temporal and spectral subtleties which contribute to its perceived qualities. It is highly encouraging that the TFS strategy used here could extract usable pulse timing information from the acoustic TFS of speech and that ITDs could be perceived even when imbedded in streams of dynamically changing rate. However, in order to study the effects of pulse timing on binaural hearing, we chose to introduce a rather large artificiality by including only ITD cues. As discussed above, the presence of ILD cues may overshadow the benefit from ITD cues in more realistic listening situations, and would certainly have complicated the determination of pulse timing utility in these experiments.

The trade-off between better speech recognition at high rates and better pulse timing sensitivity at low rates presents a conundrum, and it may be instructive to discuss the underlying mechanisms. Envelope fluctuations may be better represented by high rate pulse trains due to two factors. First, a higher pulse rate represents a higher sampling rate, so that high frequency fluctuations will be more faithfully re-presented to the auditory nerve. The other factor is less obvious. In this study, the current levels at which stimuli are comfortable were found to be about 10-20 current units lower with high rates than with low rates. The drop in thresholds, though, were usually found to be greater, around 30-40 current units. Thus, the dynamic range for high rates can be 10-30 current units larger than for low rates- a 20% to 100% increase. This larger dynamic range may be responsible for superior envelope representation and speech understanding with high-rate stimulation. It is unknown which of these two phenomena dominate the improvement in speech understanding at high rates, but the answer may have implications for any implementation of low-rate stimulation for improved spatial hearing and suggests that additional techniques, such as pulse width modulation, may be needed to play a role in improving dynamic range for better envelope representation with low rates.

The current results indicate that CI listeners' spatial hearing improves without severe detriment to speech understanding in quiet when given binaural pulse timing cues at low and mixed rates. This finding suggests that some bilateral CI listeners may benefit from the inclusion of low- and/or mixed-rate TFS-derived pulse timing. However, it is unknown how the TFS and CIS strategies presented here would compare in realistic listening conditions, where listeners are also provided ILD and spectral tilt cues. Due to the independence of bilateral processing steps, we expect that the TFS strategy described here would provide the same ITD cues if it were implemented on independent bilateral processors. Additionally, although not directly tested here, the TFS strategy should have the ability to better represent $f_0$ voicing information, which could improve speech understanding in complex listening environments.

**Conclusion**

The findings from this experiment show that pulse timing can be a useful cue for ITD lateralization and discrimination of speech sounds with bilateral CIs. Pulse timing cues here derived from acoustic TFS were observed to be most useful when obtained from a low-frequency channel, resulting in low rates, and retained significant utility when presented on four apical electrodes in conjunction with high rates on four basal electrodes. Additionally, listeners generally understood speech in quiet better when TFS-derived pulse timing cues were present, even though the TFS was not extracted from the same channels as the envelopes. Given the fact that these listeners are normally deprived of pulse timing cues representative of acoustic TFS, one would expect that their observed sensitivity to these cues might be lower here than if the listeners were routinely exposed to them. The tendencies observed here regarding differences in performance among listeners with different strategies and rate combinations suggest that future clinical devices should include processing strategies which preserve

acoustic TFS information in pulse timing and perhaps have options for different rate combination settings.

**References**

Agrawal, S (2008). Spatial hearing abilities in adults with bilateral cochlear implants. Doctor of Philosophy, University of Wisconsin-Madison.

Agrawal, S, Litovsky, R Y, Jones, G L and Van Hoesel, R J M (2006). Correlates of Precedence Effect In Bilateral Cochlear Implant Users under the conditions of Direct Stimulation and in Free-Field. 2006 Midwinter Meeting of the Association for Research in Otolaryngology, Baltimore, MD.

Algazi, V R, Duda, R O and Thompson, D M (2001). The CIPIC HRTF database. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York.

Arbogast, T L, Mason, C R and Kidd, G (2002). "The effect of spatial separation on informational and energetic masking of speech." J Acoust Soc Am 112(5): 2086-2098.

Arbogast, T L, Mason, C R and Kidd, G (2005). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners." J Acoust Soc Am 117(4): 2169-2180.

Arnoldner, C, Riss, D, Brunner, M, Durisin, M, Baumgartner, W-D and Hamzavi, J-S (2007). "Speech and music perception with the new fine structure speech coding strategy: preliminary results." Acta oto-laryngologica 127: 1298-1303.

Bronkhorst, A and Plomp, R (1988). "The effect of head-induced interaural time and level differences on speech intilligibility in noise." J Acoust Soc Am 83(4): 1508-1516.

Brughera, A, Dunai, L and Hartmann, W M (2013). "Human interaural time difference thresholds for sine tones: the high-frequency limit." J Acoust Soc Am 133(5): 2839-2855.

Carlyon, R P, Deeks, J M and McKay, C M (2010). "The upper limit of temporal pitch for cochlear-implant listeners: stimulus duration, conditioner pulses, and the number of electrodes stimulated." J Acoust Soc Am 127(3): 1469-1478.

Chatterjee, M and Robert, M E (2001). "Noise Enhances Modulation Sensitivity in Cochlear Implant Listeners: Stochastic Resonance in a Prosthetic Sensory System?" JARO - Journal of the Association for Research in Otolaryngology 2: 159-171.

Churchill, T H, Ihlefeld, A, Kan, A, Carlyon, R P and Litovsky, R Y (2012). Unilateral Versus Bilateral Temporal Fine Structure Processing in Bilateral Cochlear Implant Users. Winter Meeting of the Association for Research in Otolaryngology, San Diego, CA.

Darwin, C (2011) "Auditory Scene Analysis and Perceptual Organisation: an update."

Drennan, W R, Won, J H, Dasika, V K and Rubinstein, J T (2007). "Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise." JARO - Journal of the Association for Research in Otolaryngology 8: 373-383.

Freyman, R L, Helfer, K S, McCall, D and Clifton, R (1999). "The role of perceived spatial separation in the unmasking of speech." J Acoust Soc Am 106(6): 3578-3588.

Galvin, J J, 3rd and Fu, Q J (2005). "Effects of stimulation rate, mode and level on modulation detection by cochlear implant users." J Assoc Res Otolaryngol 6(3): 269-279.

Garadat, S N, Litovsky, R Y, Yu, G and Zeng, F G (2009). "Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech." J Acoust Soc Am 126(5): 2522-2535.

Hafter, E R, Dye, R H and Wenzel, E (1983). "Detection of interaural differences of intensity in trains of high-frequency clicks as a function of interclick interval and number." J Acoust Soc Am 73(5): 1708-1713.

Hawley, M L, Litovsky, R Y and Colburn, H S (1999). "Speech intelligibility and localization in a multi-source environment." J Acoust Soc Am 105(6): 3436-3448.

Hawley, M L, Litovsky, R Y and Culling, J F (2004). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer." J Acoust Soc Am 115(2): 833.

Ihlefeld, A and Litovsky, R Y (2012). "Interaural level differences do not suffice for restoring spatial release from masking in simulated cochlear implant listening." PLoS One 7(9): e45296.

Ihlefeld, A and Shinn-Cunningham, B (2008). "Spatial release from energetic and informational masking in a selective speech identification task." J Acoust Soc Am 123(6): 4369-4379.

Jones, G L, Litovsky, R Y and Hoesel, R V (2008). ITD Sensitivity in Electrical Hearing : Simulating Channel Interaction Effects in Listeners with Normal Hearing. 30th Midwinter Meeting of the Association for Research in Otolaryngology, Phoenix, AZ.

Kan, A, Stoelb, C, Litovsky, R Y, and Goupell, M J (2013) "Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users", J Acoust Soc Am, in press

Kiang, N Y and Moxon, E C (1972). "Physiological considerations in artificial stimulation of the inner ear." Ann Otol 81(5): 714-730.

Kiang, N Y, Moxon, E C and Levine, R A (1970). Auditory Nerve Activity in Cats with Normal and Abnormal Cochleas, Massachusetts Institute of Technology and Harvard Medical School.

Kidd, G, Arbogast, T L, Mason, C R and Gallun, F J (2005). "The advantage of knowing where to listen." J Acoust Soc Am 118(6): 3804-3815.

Kidd, G, Jr., Best, V and Mason, C R (2008). "Listening to every other word: examining the strength of linkage variables in forming streams of speech." J Acoust Soc Am 124(6): 3793-3802.

Kidd, G, Mason, C R, Best, V and Marrone, N (2010). "Stimulus factors influencing spatial release from speech-on-speech masking." J Acoust Soc Am 128(4): 1965-1978.

Kidd, G, Mason, C R, Deliwala, P S, Woods, W S and Colburn, H S (1994). "Reducing informational masking by sound segregation." J Acoust Soc Am 95(6): 3475-3480.

Kidd, G, Mason, C R and Gallun, F J (2005). "Combining energetic and informational masking for speech identification." J Acoust Soc Am 118(2): 982-992.

Kidd, G, Mason, C R, Rohtla, T L and Deliwala, P S (1998). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns." J Acoust Soc Am 104(1): 422-431.

Laback, B and Majdak, P (2008). "Binaural jitter improves interaural time-difference sensitivity of cochlear implantees at high pulse rates." PNAS 2007(2): 814-817.

Laback, B, Majdak, P, Schmid, K and Baumgarntner, W-d (2005). Sensitivity to Interaural Time Delay in Stimulus Fine Structure and Onset / Offset of a Bilateral Cochlear Implant Listener. 28th Midwinter Research Meeting of the Association for Research in Otolaryngology: 1-2.

Laback, B, Pok, S-M, Baumgartner, W-D, Deutsch, W a and Schmid, K (2004). "Sensitivity to Interaural Level and Envelope Time Differences of Two Bilateral Cochlear Implant Listeners Using Clinical Sound Processors." Ear Hear 25(5): 488-500.

Litovsky, R, Parkinson, A, Arcaroli, J and Sammeth, C (2006). "Simultaneous bilateral cochlear implantation in adults: a multicenter clinical study." Ear Hear 27(6): 714-731.

Litovsky, R Y, Jones, G L, Agrawal, S and van Hoesel, R (2010). "Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans." J Acoust Soc Am 127(1): 400-414.

Litovsky, R Y, Parkinson, A and Arcaroli, J (2009). "Spatial hearing and speech intelligibility in bilateral cochlear implant users." Ear Hear 30(4): 419-431.

Loizou, P C, Hu, Y, Litovsky, R, Yu, G, Peters, R, Lake, J and Roland, P (2009). "Speech recognition by bilateral cochlear implant users in a cocktail-party setting." J Acoust Soc Am 125(1): 372-383.

Loizou, P C, Hu, Y, Litovsky, R, Yu, G, Peters, R, Lake, J and Roland, P (2009). "Speech recognition by bilateral cochlear implant users in a cocktail-party setting." The Journal of the Acoustical Society of America 125: 372-383.

Loizou, P C, Poroy, O and Dorman, M (2000). "The effect of parametric variations of cochlear implant processors on speech understanding." The Journal of the Acoustical Society of America 108: 790-802.

Long, C J, Carlyon, R P, Litovsky, R Y and Downs, D H (2006). "Binaural unmasking with bilateral cochlear implants." JARO - Journal of the Association for Research in Otolaryngology 7: 352-360.

Macpherson, E A and Middlebrooks, J C (2002). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited." J Acoust Soc Am 111(5): 2219.

Majdak, P, Laback, B and Baumgartner, W-D (2006). "Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing." The Journal of the Acoustical Society of America 120: 2190-2201.

Misurelli, S M and Litovsky, R Y (2012). "Spatial release from masking in children with normal hearing and with bilateral cochlear implants: effect of interferer asymmetry." J Acoust Soc Am 132(1): 380-391.

Moore, B C J (2003). "Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants." Otol Neurotol 24: 243-254.

Morse, R P and Meyer, G F (2000). "The practical use of noise to improve speech coding by analogue cochlear implants." Chaos, Solitons, and Fractals 11: 1885-1894.

Nie, K, Stickney, G and Zeng, F-G (2005). "Encoding frequency modulation to improve cochlear implant performance in noise." IEEE transactions on bio-medical engineering 52(1): 64-73.

Poon, B B, Eddington, D K, Noel, V and Colburn, H S (2009). "Sensitivity to interaural time difference with bilateral cochlear implants : Development over time and effect of interaural electrode spacing." J Acoust Soc Am 126(2): 806-905.

Riss, D, Arnoldner, C, Baumgartner, W-D, Kaider, A and Hamzavi, J-S (2008). "A new fine structure speech coding strategy: speech perception at a reduced number of channels." Otol Neurotol 29: 784-788.

Riss, D, Arnoldner, C, Reiss, S, Baumgartner, W-D and Hamzavi, J-S (2009). "1-year results using the Opus speech processor with the fine structure speech coding strategy." Acta oto-laryngologica 129: 988-991.

Riss, D, Hamzavi, J-S, Katzinger, M, Baumgartner, W-D, Kaider, A, Gstoettner, W and Arnoldner, C (2011). "Effects of fine structure and extended low frequencies in pediatric cochlear implant recipients." International journal of pediatric otorhinolaryngology 75: 573-578.

Rubinstein, J T and Miller, C a (1999). "How do cochlear prostheses work?" Current opinion in neurobiology 9: 399-404.

Rubinstein, J T Y, Wilson, B S, Finley, C C and Abbas, P J (1999). "Pseudospontaneous activity : stochastic independence of auditory nerve ¢ bers with electrical stimulation." Hear Res 127: 108-118.

Schatzer, R, Krenmayr, A, Au, D K K, Kals, M and Zierhofer, C (2010). "Temporal fine structure in cochlear implants: preliminary speech perception results in Cantonese-speaking implant users." Acta oto-laryngologica 130: 1031-1039.

Schleich, P, Nopp, P and D'Haese, P (2004). "Head Shadow, Squelch, and Summation Effects in Bilateral Users of the MED-EL COMBI 40/40+ Cochlear Implant." Ear Hear 25(3): 197-204.

Shannon, R V, Zeng, F-G, Kamath, V, Wygonski, J and Ekelid, M (1995). "Speech recognition with primarily temporal cues." Science 270(5234): 303-304.

Sit, J-j (2007). An Asynchronous , Low-Power Architecture for Interleaved Neural Stimulation , using Envelope and Phase Information. Doctor of Philosophy, Massachusetts Institute of Technology.

Sit, J-j, Simonson, A M, Oxenham, A J, Faltys, M A and Sarpeshkar, R (2007). "A Low-Power Asynchronous Interleaved Sampling Algorithm for Cochlear Implants That Encodes Envelope and Phase Information." IEEE Transactions on Biomedical Engineering 54(1): 138-149.

Smith, R (1979). "Adaptation, saturation, and physiological masking in single auditory-nerve fibers." J Acoust Soc Am 65(1): 166-178.

Smith, Z M and Delgutte, B (2008). "Sensitivity of Inferior Colliculus Neurons to Interaural Time Differences in the Envelope Versus the Fine Structure With Bilateral Cochlear Implants." J Neurophysiol 99: 2390-2407.

Smith, Z M, Delgutte, B and Oxenham, A J (2002). "Chimaeric sounds reveal dichotomies in auditory perception." Nature 416: 87-90.

Studebaker, G A (1985). "A "rationalized" arcsine transform." J Speech Hear Res 28: 455 - 462.

van Hoesel, R J M (2007). Peak-derived timing stimulation strategy for a multichannel cochlear implant. United States.

van Hoesel, R J M, Böhm, M, Pesch, J, Vandali, A, Battmer, R D and Lenarz, T (2008). "Binaural speech unmasking and localization in noise with bilateral cochlear implants using envelope and fine-timing." J Acoust Soc Am 123(4): 2249-2263.

van Hoesel, R J M and Clark, G M (1997). "Psychophysical studies with two binaural cochlear implant subjects." J Acoust Soc Am 102(1).

van Hoesel, R J M, Jones, G L and Litovsky, R Y (2009). "Interaural time-delay sensitivity in bilateral cochlear implant users: effects of pulse rate, modulation rate, and place of stimulation." JARO - Journal of the Association for Research in Otolaryngology 10: 557-567.

van Hoesel, R J M, Ramsden, R and Odriscoll, M (2002). "Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user." Ear Hear 23: 137-149.

van Hoesel, R J M and Tyler, R S (2003). "Speech perception, localization, and lateralization with bilateral cochlear implants." J Acoust Soc Am 113(3): 1617-1630.

Vermeire, K, Punte, K and Heyning, V D (2010). "Better Speech Recognition in Noise with the Fine Structure Processing." ORL 72: 305-311.

Wichman, F and Hill, N (2001). "The psychometric function: I. Fitting, sampling, and goodness of fit." Perception & psychophysics 63(8): 1293-1313.

Wightman, F and Kistler, D (1992). "The dominant role of low-frequency interaural time differences in sound localization." J Acoust Soc Am 91(3): 1648-1661.

Wilson, B, Finley, C, Lawson, D, Wolford, R, Eddington, D K and Rabinowitz, W (1991). "Better speech recognition with cochlear implants." Nature 352(236-238).

Wilson, B, Sun, X, Schatzer, R and Wolford, R (2004). "Representation of fine structure or fine frequency information with cochlear implants." International Congress Series 1273: 3-6.

Zierhofer, C (2003). Electrical Nerve Stimulation based on Channel Specific Sampling Sequences. United States.

# Chapter 5

# Discussion and Conclusions

The studies reported in this dissertation have explored both electric and normal hearing (NH) mechanisms and strategies to enhance hearing with cochlear implants (CIs). The results of these studies have contributed to a better understanding of those aspects of the peripheral auditory system that support speech understanding and binaural hearing, they have suggested vocoder parameters that may simulate certain aspects of electric hearing, and they have shown that redundant, low-rate, TFS-based pulse timing supports improved spatial hearing for bilateral CI listeners.

The sine-phase harmonic carrier ("H0") used in the vocoder study of Chapter 2 produced response patterns on simulated auditory nerve (AN) fibers that were highly localized in time for neurons of characteristics frequency (CF) greater than 1.5kHz (see figure 7 of that chapter). This type of temporal response pattern is typical of electric hearing, where ANs fire synchronously with electrical stimulation. Whereas activation delays between fibers of different CFs are truly lost in electric hearing, this effect cannot easily be reproduced in NH, even with a vocoder. However, the synchronicity of neural response

to the speech vocoded with the H0 carrier is evident, and the resulting speech recognition scores are lower than with other carriers. In addition, stimuli vocoded with the H0 carrier produced the lowest average neural envelope correlation coefficients, indicating they were the poorest at carrying speech temporal envelopes. The simulated neural response patterns due to sine- and noise-vocoded speech stimuli show much less synchronous response and produce much higher neural envelope correlation coefficients. These findings yield two important implications. First, the sine-phase harmonic complex carrier H0 provides a better vocoder simulation of electric hearing than either sine tones or white noise. The temporal characteristics of AN response to electrical stimulation have been implicated in the loss of some sensitivities associated with NH, and it behooves us to reproduce the most realistic simulations of electric hearing in research that purports to simulate CI hearing with a vocoder.

Second, the worse performance and lower neural envelope correlation coefficients with carrier H0 suggest that *better* speech recognition (presumably due to better neural envelope representations) may be produced by desynchronizing neural response. Indeed, the improved results with random-phase harmonic carriers support this conclusion. There are several ways that this finding could be applied to CIs. In general, we conclude that better speech recognition might result if neurons could fire independently, like they do via NH mechanisms. The implementation of interleaved electrode stimulation, such as exists on all contemporary CIs, is a good first step towards this goal. By disallowing the simultaneous activation of multiple electrodes, the temporal synchronicity of neural activation is greatly reduced. Another step taken in the evolution of CIs has been the move to higher rates of pulsatile stimulation. It has been observed that higher stimulation rates result in better speech understanding and require lower current levels to achieve equal loudness percepts (see Chapter 4). If high current levels are associated with more current spread and synchronous activation, perhaps the lower current levels required by high rates of stimulation improve neural desynchronization. Recent

efforts to better focus the spatial extent of neural activation by current steering or multipolar stimulation could result in additional desynchronization of neural response. Further reduction in synchronous activation will likely require a better understanding of the nature of neural activation by electric stimulation and the path of current through the inner ear. Candidates for evaluation for more selective stimulation include pulse width modulation techniques and novel waveforms that exploit neurons' biophysical properties.

In order to produce CIs that provide the best possible inputs at the auditory periphery, we not only need an exquisite understanding of how electrical stimulation activates the AN, but also an understanding of how AN activation patterns are processed more centrally. In the study reported in Chapter 3, we explored several potential processing mechanisms for binaural spatial hearing by examining the ability of binaural correlation metrics extracted from AN simulations to predict stimulus azimuths. Because activation delays due to the cochlear traveling wave are absent with CIs, we tested both the case of internal delay lines with as-yet undiscovered associated anatomy and the case of cochlear delays only. Additionally, since it is thought that interaural time differences (ITDs) in the envelopes and temporal fine structure (TFS) may be processed by different neurophysiological machinery, by decomposing the cross-correlation functions into phase-sensitive ("difcor") and phase-insensitive ("sumcor") components, we sought to learn the respective contributions of envelope and TFS features to lateralization. The results indicated several interesting findings.

One of the main findings of the study was that there were few differences between the azimuth-predictive abilities of the internal delay line model, as estimated by cross-correlation peak abscissae, and the stereausis model, as estimated by binaural correlation plane diagonal sum peak abscissae. The internal delay line models often produced fair estimates of broadband signals' azimuths

and were generally better than stereausis models at these predictions. This finding suggests that cross-CF coincidence detection using AN inputs alone may be a weak mechanism for binaural hearing. With current spread in CIs, the stereausis model using AN activation patterns would be completely unreliable for the calculation of binaural cues because of the assumed completely synchronous activation of all AN fibers affected by the current from a given electrode. In the CI case, there would be no usable pattern among the across-CF binaural correlations because temporal activation patterns of a fiber of a given CF in one ear would correlate equally well to those of fibers of many CFs in the other ear, regardless of ITD. The spatial control of cochlear current paths has not yet reached the resolution necessary to directly test stereausis-driven binaural calculations with electric hearing. However, the fact that many bilateral CI users are sensitive to ITDs in pulse timing seems to further advocate for the presence of an internal delay line mechanism. It is entirely possible that central transformations of the AN patterns render them more appropriate as inputs to a stereausis calculation and that these transformations may remain partially effective even for AN patterns generated by electrical stimulation. For example, broad or local inhibition networks at the level of the cochlear nucleus (CN) may still be active. These hypotheses eagerly await the development of a model of CN response to AN patterns produced by electrical stimulation.

The stimuli tested included the sine-phase and random-phase harmonic complex carriers from the vocoder study reported in Chapter 2 (H0 and H360, respectively). When only interaural level differences (ILDs) were present in the acoustic signal, most ITD metrics were anticorrelated with azimuth for H0, while ITD metrics were positively correlated with azimuth for H360. This finding is especially interesting when we consider that the H0 carrier may be a realistic simulation of CI stimulation and that ILD cues are currently the only location cue available with CIs. Specifically, the finding suggests that binaural timing processing mechanisms may produce false lateral position

estimates when ILDs are presented alone with CIs. Furthermore, when full HRTF cues were present in the stimuli, all ITD metrics were positively correlated with azimuth for H0. Taken together, these findings strongly suggest that both ILDs and ITDs should be represented in CI stimulation in order to provide the best possible cue set for spatial hearing.

For the study reported in Chapter 4, we investigated a way to deliver accurate and accessible ITD cues with a novel CI processing strategy. The results showed that low-rate TFS-timed pulses can greatly enhance ITD sensitivity for bilaterally-implanted CI listeners. At low and mixed stimulation rates, listeners had significantly better ITD lateralization and discrimination abilities as well as better speech understanding with the TFS strategy than with the CIS strategy. These findings strongly advocate for the inclusion of low- and/or mixed-rate TFS-timed pulse strategies among those available to bilateral CI users. They also indicate that the stereausis mechanism is unlikely to be the exclusive calculator for binaural timing disparities. Were that the case, along with exquisite timing control, much sharper control of spatial activation patterns would be necessary to produce the required binaural correlation patterns. We have successfully shown that ITD sensitivity to stereo speech signals is available with the current state of the art. However, we are still far from a full understanding of the pathological effects of electric stimulation on binaural calculators, and much study of the question remains to be done.

While the findings of Chapter 2 indicated that cross-fiber, within-channel temporal synchrony of neural response may be pathologic to speech understanding in noise, *across-channel* temporal envelope synchrony may enhance it (Crouzet & Ainsworth, 2001). If neural envelopes may be recovered from TFS, and cross-channel envelope correlation aids in speech understanding, it may be hypothesized that cross-channel TFS correlation could aid in speech understanding. It is not trivial, however, to reproduce the same TFS signal at different locations in the cochlea with normal hearing. In Chapter 4, we showed that

redundant, TFS-derived pulse timing resulted in superior speech recognition for low and mixed stimulation rates. Because the pulse timing with the CIS strategy also presented redundant information, we may conclude that it was the relation of the pulse timing information to the speech waveform with the TFS strategy that resulted in the improvement. If neural envelopes are able to be reconstructed from information in TFS-timed pulses, this yet another reason to carry acoustic information in the pulse timing of CIs.

Contemporary efforts to advance CI technology are rightly centered on improving the devices' temporal and spatial/spectral resolution. It may seem curious that we might improve temporal resolution by improving spectral resolution with CIs- this approach seems to fly in the face of the uncertainty principle. However, the layout, function, and nonlinearities of the auditory periphery may free us from such chains up to a point (Oppenheim & Magnasco, 2013). Acoustic attributes can translate into neural parameters such as place of activation, firing rate, and periodicity of activation. Additionally, AN fibers of different spontaneous rate (SR) and threshold may be responsible for the encoding of different acoustic attributes. Furthermore, nonlinear dependencies, such as the level-dependence of phase-locking phase, could introduce additional cues to acoustic information. Psychoacousticians (and CI signal processing algorithms) commonly utilize the decomposition of acoustic signals in the time domain into envelopes and TFS carriers, a procedure known as "modulation filtering." It has been found that envelope and TFS often carry distinct cues for various hearing percepts and abilities, and it is thus assumed that the peripheral auditory system must somehow encode envelopes and TFS information. However, this general assumption is not entirely substantiated and recent studies have called it into question.

One of the most commonly-used methods of envelope extraction by modulation filtering, the Hilbert transform, relies on the assumption that temporal envelope fluctuations are of much lower frequency than the bandwidth of the original signal (Hartmann, 2004; Schimmel & Atlas, 2005). This condition may not be met when the Hilbert transform is used for envelope extraction in vocoder or CI processing. Additionally, it has been demonstrated that envelopes may be recovered from TFS at the output of auditory filters (Apoux, Millman, Viemeister, Brown, & Bacon, 2011; Ghitza, 2001) and that these recovered envelopes carry the information necessary for speech understanding (Gilbert & Lorenzi, 2006; Sheft, Ardoint, & Lorenzi, 2008). These conclusions extend also to vocoded speech (Shamma & Lorenzi, 2013). The apparent failure of modulation filtering to create in temporal envelopes and TFS an orthogonal basis for information at the output of the auditory filters and thus for information in the patterns of AN activation suggests future experiments should evaluate other decomposition techniques.

The decomposition of cross-correlations of neural response into phase-sensitive and phase-insensitive portions via difcor and sumcor calculations, respectively, represents a prescient step towards a fuller exploration of the auditory code space. If we take x to be a 2×N matrix whose rows are binary neural “raster” responses or post-stimulus time histograms (PSTHs) to the original and inverted waveforms, then the envelope and TFS response can be represented in the rows of y=Ax, where

$$A = \begin{bmatrix} 0.5 & 0.5 \\ 1 & -1 \end{bmatrix}$$

Chapter 3 contained a suggestion for further improving the phase-insensitive response representation, namely, to include a multiplicity of signal phases in the average vice just two. Using an M×N raster or PSTH set x due to M waveforms with phase angles

$$\theta = 0 \; to \; 2\pi \left( \frac{M-1}{M} \right)$$

generalizes A to a 2×M matrix such as

$$A = \begin{bmatrix} 1/M & 1/M & 1/M & \dots & 1/M & 1/M \\ \cos\left(\frac{2\pi(0)}{M}\right) & \cos\left(\frac{2\pi(1)}{M}\right) & \cos\left(\frac{2\pi(2)}{M}\right) & \dots & \cos\left(\frac{2\pi(M-2)}{M}\right) & \cos\left(\frac{2\pi(M-1)}{M}\right) \end{bmatrix}$$

We now conclude with an example of such a decomposition. The input stimulus, the single word "goose" spoken in quiet, was rotated in phase by $2\pi$ in increments of $\pi/8$ using the Hilbert transform (M=16). Each of the 16 rotated speech tokens was used to generate simulated inner hair cell (IHC) voltages (deterministic) and AN response PSTHs (stochastic) for each of 100 IHCs and AN fibers with characteristic frequencies (CFs) logarithmically spaced between 100 Hz and 10 kHz. For both IHC and AN simulations, the 16 outputs for each CF were represented by row vectors in a matrix, x, which was multiplied by the 2×16 matrix A, yielding the two-row matrix y: y=Ax. The rows of y represent phase-insensitive and phase-sensitive parts of the response patterns, which are shown, for all CFs, for a voiced segment of the stimulus token in figure 1. The speech token could be "reconstructed" from the IHC response patterns by summing the resulting voltages along the CF dimension. Only speech reconstructed from the phase-sensitive IHC response was intelligible. A speech-reconstruction from the AN patterns using a set of 100 complex-exponential impulse responses also yielded an intelligible word only for the TFS decomposition. While beyond the scope of this dissertation, these extended decompositions provide an excellent basis for future research.

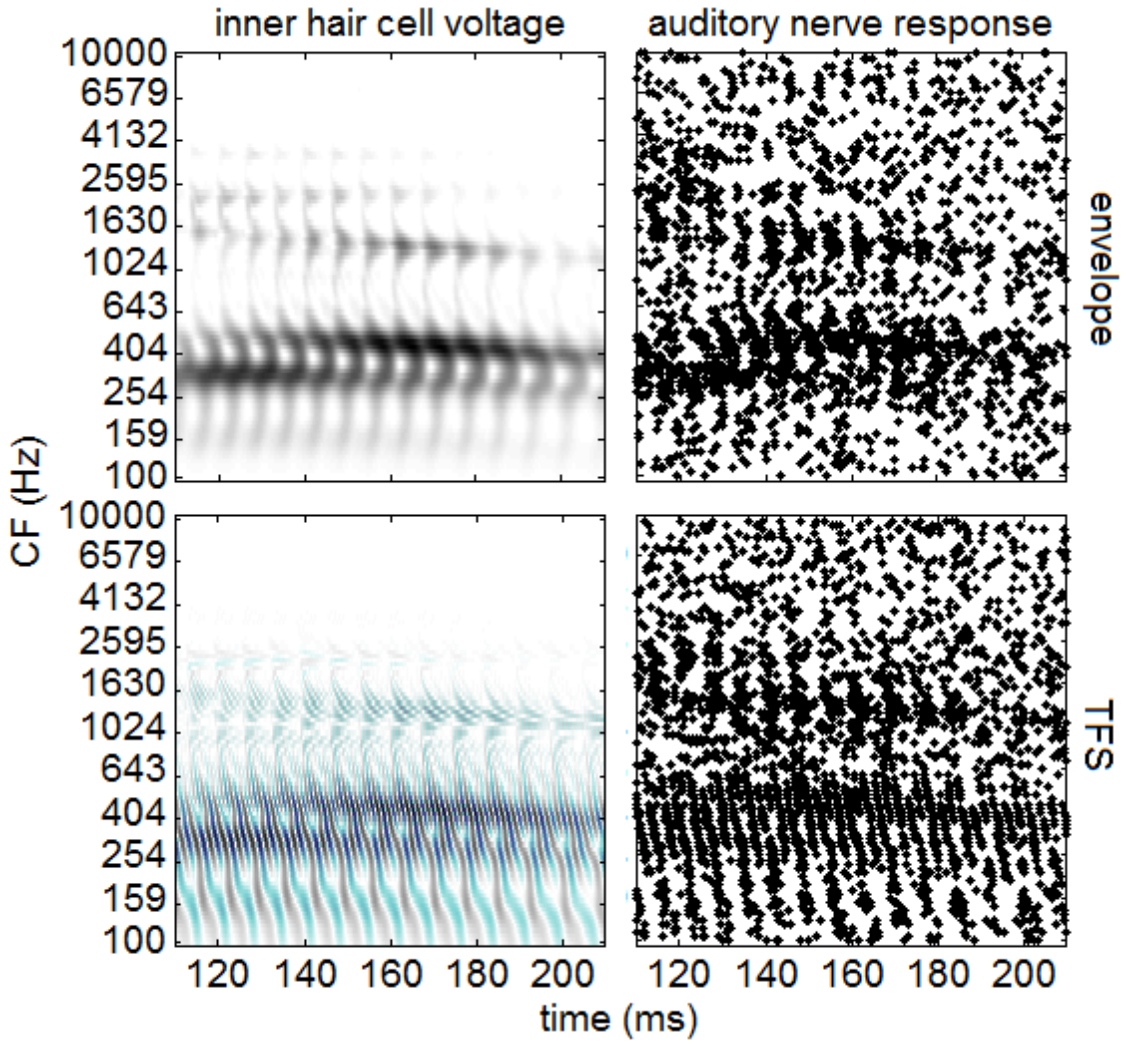



Figure 1. Phase insensitive (envelope) and phase-sensitive (TFS) decomposition of simulated IHC voltage and AN response patterns for a voiced segment of the word "goose" spoken in quiet. For the $V_{IHC}$ patterns, response strength is represented by grayscale intensity and plotted as a function of time and cell CF. Blue portions indicate negative values. For the AN patterns, a high spontaneous rate fiber simulation was used and histogram counts above a given threshold (chosen for image clarity) are represented by black dots.

**References**

Apoux, F., Millman, R. E., Viemeister, N. F., Brown, C. a, & Bacon, S. P. (2011). On the mechanisms involved in the recovery of envelope information from temporal fine structure. The Journal of the Acoustical Society of America, 130(1), 273. doi:10.1121/1.3596463

Crouzet, O., & Ainsworth, W. A. (2001). On the Various Influences of Envelope Information on the Perception of Speech in Adverse Conditions: An Analysis of Between-Channel Envelope Correlation.

Ghitza, O. (2001). On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. The Journal of the Acoustical Society of America, 110(3), 1628. doi:10.1121/1.1396325

Gilbert, G., & Lorenzi, C. (2006). The ability of listeners to use recovered envelope cues from speech fine structure. The Journal of the Acoustical Society of America, 119(4), 2438. doi:10.1121/1.2173522

Hartmann, W. M. (2004). Signals, Sound, and Sensation. American Institute of Physics.

Oppenheim, J. N., & Magnasco, M. O. (2013). Human Time-Frequency Acuity Beats the Fourier Uncertainty Principle. Physical Review Letters, 110(4), 044301. doi:10.1103/PhysRevLett.110.044301

Schimmel, S., & Atlas, L. (2005). Coherent Envelope Detection for Modulation Filtering of Speech. Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., 1(7), 221–224. doi:10.1109/ICASSP.2005.1415090

Shamma, S., & Lorenzi, C. (2013). On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system, 133(May). doi:10.1121/1.4795783

Sheft, S., Ardoint, M., & Lorenzi, C. (2008). Speech identification based on temporal fine structure cues. J Acoust Soc Am, 124(1), 562–575. doi:10.1121/1.2918540