NORTHWESTERN UNIVERSITY

Towards Optical Cochlear Implants

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Communication Sciences and Disorders

By

Yingyue Xu

Evanston, IL

June 2019

Abstract

Towards Optical Cochlear Implants

Yingyue Xu

This thesis focuses on the development of a cochlear implant (CI) that uses photons to stimulate surviving auditory neurons in severe-to-profoundly deaf individuals. The benefit of optical over electrical stimulation is its spatial selectivity with the potential to create significantly more independent channels to encode acoustic information and likely enhances the CI users' performance in challenging listening environments. Towards the development of an optical cochlear implant, there are two challenges that we addressed in the dissertation: firstly, the development of the light delivery system that can be implanted into the cochlea and evoke auditory responses; secondly, the creation of a coding strategy with temporal periodic cues to accommodate for the low rate of infrared neural stimulation (INS).

INS has demonstrated superior potential in more selective stimulation of the cochlea using single flat polished optical fibers. To enable the delivery of photons to the cochlea for future clinical applications, more sophisticated light delivery systems are required. Aim 1 of the thesis is to develop and test such systems. In this dissertation, we have developed a light delivery system using small light sources that evoked auditory responses in vivo.

INS is limited by the rate of stimulation. A low rate speech coding strategy is needed to enable this technology for cochlear implants. A novel speech coding strategy has been developed in our lab that fits such requirements, which elicited speech perception in CI users. Such coding

strategy encoded the temporal cues of speech (F_0 and its harmonics). Aim 2 of the thesis is to examine the functional role of speech temporal cues and their neural representations to give guidance for proper representations of speech temporal cues in CIs. This dissertation demonstrated that the temporal cues within 500 Hz are essential for speech perception in normal hearing listeners under both quiet and noisy listening environment. The dissertation also uncovered the neural processing of temporal cues of speech.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The auditory system is part of the human sensory systems that transduces, processes and integrates acoustic information. Individuals with hearing losses of different extents face various challenges in life and might need constant monitoring of hearing thresholds and restoration of the access to sounds if necessary. For individuals with severe-to-profound hearing loss, the state of art solution for hearing restoration lies in cochlear implants (CIs). Although CIs are considered unparalleled successful neural prostheses on the market, the patient outcomes vary largely and show CI users are challenged in noisy environments and music perception [1, 2]. One of the two obstacles CI is the limited number of independent channels and the lack of temporal cues. This dissertation addresses these two problems in Chapter 2 and 3 and 4. The results are summarized in Chapter 5.

1.1 The anatomy and physiology of the peripheral auditory system

The peripheral auditory system consists of three major components: the outer ear, the middle ear, and the inner ear. Here is a brief introduction of the anatomy and physiology of these three parts. [3]

The first part is the outer ear, which consists of the pinna and the outer ear canal. The pinna is mainly cartilaginous and contains many convoluted folds, which captures and transmits sound through the external acoustic meatus to the eardrum. The ear canal has a peripheral segment consists of cartilaginous walls and a medial segment of bony walls.

The second part is the middle ear, which consists of the tympanic membrane, the middle ear ossicles, the middle ear muscles, and the Eustachian tube. The middle ear is an air-filled cavity either surrounded by the temporal bone in higher mammals or by the auditory bulla, which is a closed cavity except for the Eustachian tube that links to the nasopharynx. The middle ear consists of four main components: the tympanic membrane, ossicular chain made of the malleus, incus, and the stapes, the middle ear muscles, ligaments, and the Eustachian tube. The tympanic membrane is a cone-shaped structure on the lateral border of the middle ear. The tympanic membrane connects with the manubrium of the malleus over the length of its full radius 12 o'clock. The malleus connects with the incus. The incus is connected with the stapes through the incudostapedial joint. The footplate of the stapes sits on the oval window of the cochlea. The chain of middle ear ossicles transmits the vibration of the tympanic membrane to the cochlea. The two middle ear muscles are the stapedius and tensor tympani muscles, which can stiffen the ossicular chain when activated. The stapedius muscle is located in a bony canal in the posterior wall of the tympanic cavity and is connected to the stapes through its protruding tendon from the canal. The tensor tympani muscle is located in a bony canal parallel to the Eustachian tube and is connected to the manubrium of the malleus by the tendon. The two middle ear muscles are striated muscles. The middle ear muscles contract in response to in high-intensity sound stimuli, known as the acoustic reflex [4]. Under the activation of the acoustic reflex, the tensor tympani muscle pulls the malleus inwards and subsequently cases an inward motion of the tympanic membrane, whereas the stapedius muscle pulls the stapes outward away from the oval window. Therefore, the two opposite pulling direction of the stapedius and tensor tympani can stiffen the ossicular chain, resulting in a reduction in the sound transmission in the middle ear. The effects of acoustic reflex

are two folds: (1) it decreases the energy transmitted from the out ear to the cochlea; (2) it alters the pressure that is generated in the cochlea and transmitted into the ear canal, such as otoacoustic emissions (OAE). Thus, middle ear muscle contractions can act not only on the energy absorbed by the ear by changing the reflectance of the middle ear but also on the reverse transmission of OAE signals [5, 6].

The third part is the inner ear, which consists of the cochlea and the vestibular system. The cochlea is a snail-shaped structure that contains two to five complete turns in mammals [7]. The cochlea is embedded in a bony structure (otic capsule) within the temporal bone or the wall of the bulla [7]. The bony cochlea is coiled around the modiolus, its bony central core. The inner membranous portion of the cochlea contains three adjacent ducts, scala vestibuli, scala media, and scala tympani, through most of the coil length which are separated by two membranes: the Reissner's membrane and the basilar membrane. Reissner's membrane separates scala vestibuli, scala media, whereas the basilar membrane separates the scala media and scala tympani. Scala tympani and scala vestibuli are joined via the helicotrema at the apex of the cochlea, the tip of the coil. Both scala tympani and scala vestibuli are filled with perilymph, which has a similar ionic composition to the cerebrospinal fluid. Scala media is between scala tympani and scala vestibuli, which ends near the apex of the cochlea and is filled with high potassium concentration endolymph. There are two openings at the base of the cochlea: the oval window of scala vestibuli, which holds the stapes footplate and the round window of scala tympani, which is covered by the flexible round window membrane. The vibration of the stapes displaces the fluid in scala vestibuli and excites a traveling wave on the BM. The peak of this traveling wave occurs at different sites along the cochlear partition for different frequencies. Low frequencies peak near the apex and high

frequencies near the base. This site of maximal vibration is determined by the mechanical properties of the BM, which is narrow and stiff near the base of the cochlea and wide at the apex. On top of the BM sits the organ of Corti, which is also smaller at the base than at the apex of the cochlea. The sensory cells of auditory systems, the hair cells, are part of the organ of Corti and are covered by the tectorial membrane. There are two types of hair cells, the inner hair cells and outer hair cells, that are distinct in terms of morphologies, spatial locations, innervations, and functions. The inner hair cells (IHCs) are flask-shaped and configured in one single row. The IHCs innervate with type I spiral ganglion cells, which are the main the cochlear afferent inputs. The outer hair cells (OHCs) are cylindrically-shaped and configured in 3 adjacent rows. The OHCs are innervated by the medial olivocochlear neurons, receiving the auditory efferent modulations. Upon the occurrence of a traveling wave, a shearing force between the tectorial membrane and organ of Corti deflects the stereocilia, the hair-like projections on top of the HCs. The deflection of the stereocilia activate the mechanically gated cation channels, allowing an inward potassium current and depolarizes the HCs. The depolarization of the IHCs leads to the excitation of the auditory nerve fibers and initiate the ascending pathway to the auditory cortex.

1.2 Hearing loss and hearing restoration

1.2.1 Hearing loss and evaluation

Hearing loss refers to an impairment of the auditory system, which can be generally classified as conductive or sensorineural. Conductive hearing loss is caused by changes of the outer and middle ear that interfere with the sound transmission to the cochlea. Sensorineural hearing loss

is typically caused by the damage of the inner ear structures, the auditory nerve, or cortical centers involved in hearing. In some cases, conductive and sensorineural hearing loss can both be observed in one patient. The hearing is assessed in the clinic or research laboratories through pure tone audiometry, electrophysiological audiometry, otoacoustic audiometry, speech perception tests, tympanometry, etc.

1.2.1.1 Pure tone audiometry

Hearing is frequently assessed by measuring the hearing acuity for different frequencies, i.e. pure tone audiometry. In routine clinical hearing tests, the range of test frequencies covers 125 – 8000 Hz. In research settings, the range of frequencies is typically wider for humans. The frequency range used to test cochlear function in animals depends on the animal species. During the pure tone audiometry, one pure tone is delivered at one time to the subject via headphone (air-conducted) or a vibrator on the temporal bone (bone-conducted). The subject is asked to respond to whether she or he can hear the pure tone. The softest pure tone that evoked 50% accurate audibility can be determined as the pure tone threshold. The sound intensity is usually measured in decibel hearing level (dB HL) in clinics, which is the sounds pressure level relative to the threshold for a young healthy individual. Thresholds between -10 and +15 dB HL are considered in the normal range, whereas threshold elevations above 15 dB HL are considered as hearing impaired. More specifically, thresholds for light hearing loss are 16 to 25 dB HL, thresholds for mild hearing loss are 26 to 40 dB HL, thresholds for moderate hearing loss are 41 to 55 dB HL,

thresholds for moderate-to-severe hearing loss are 56 to 70 dB HL, thresholds for severe hearing loss are 71 to 90 dB HL, thresholds for profound hearing loss are above 91 dB HL [8].

1.2.1.2 The electrophysiologic audiometry

The auditory function can also be assessed via objective measures such as auditory brainstem responses (ABRs) and OAEs. These tests can be conducted when supporting measures are need and/or behavior tests such as pure tone audiometry cannot be performed, such as newborns. ABRs are electrical potentials recorded from the scalp, during the delivery of acoustic clicks or pure tones to the outer ear canal. ABRs are the response of a large population of auditory nerve fibers and its size and duration reflect the synchronization of auditory neural responses to the onset of the stimuli [9]. ABRs are usually recorded with three electrodes: an active electrode, a reference electrode, and a ground electrode. For human subjects, a vertical montage: the active electrode on the forehead, the reference electrode on the mastoids or the earlobes, and the ground electrode on the low forehead. For animals, the active electrode is placed behind the pinna close to the bulla, the reference electrode is placed on the top of the skull, and the ground electrode on the low back. The ABR waveforms consist five peaks, wave I–V, which can be used to interpret the function of different sites along the auditory pathway from the cochlea to the brainstem based on the origins of the five waves. It has been suggested that wave I originates from the auditory nerve, wave II from the cochlear nucleus, wave III from the superior olivary complex, and wave IV–V from the lateral lemniscus and inferior colliculus.

1.2.2 Hearing restoration with cochlear implants

1.2.2.1 Cochlear implant technology

Cochleae of patients who are severe-to-profoundly deaf show massive losses of their hair cells and auditory neurons. Therefore, the transduction of sound induced vibrations of the basilar membrane into a series of action potentials is no longer possible. However, the remaining auditory neurons can be directly stimulated with electrical current delivered with cochlear implants (CIs) and some of the hearing can be restored. Today, CIs are the most successful devices among existing neural prostheses. About 350,000 individuals with severe-to-profound hearing loss have received CIs. Every year, approximately 50,000 severe-to-profoundly deaf receive a cochlear implant. Cochlear implants bypass the HCs and directly stimulate surviving auditory neurons for hearing restorations. There are four essentials components for current clinically-available cochlear implants: a microphone, a speech processor, a transmission link and a stimulating electrode [10]. The microphone is the input to the speech processor and captures the acoustic signal. The speech processor divides the acoustic signal into frequency bands, calculates the energy changes in each frequency band over time, and uses those values to modulate the amplitude of trains of electrical current pulses delivered to the cochlea. This amplitude information is transmitted via a radio frequency (RF) link to the implanted receiver, which is a pulsed current source. The current pulses are then delivered to different sites along the cochlear spiral with a wire bundle, the cochlear implant electrode, which was surgically inserted into scala tympani of the cochlea.

1.2.2.2 Speech processing strategies

The speech processor extracts and encodes the relevant acoustic information. Coding strategies of the first implants provided the patients with word and limited open-set speech recognition. Historically speech processing strategies started stimulating with sinusoidal currents: simultaneous analog stimulation (SAS) [11, 12]. The input for the SAS strategy is filtered into contiguous bands directed to corresponding electrode contacts. The mapping the dynamic ranges from natural hearing to electrically evoked haring is achieved by controlling the gain of the amplifiers for different bands. Moreover, the SAS strategy does not interleave the stimulation across different electrode contact. The results were not bad for a single channel. However, stimulating at neighboring channels resulted was detrimental: overlapping current fields resulted in loudness summation which was not liked by the patient [13]. Next, pulsatile stimuli were used instead of sinusoidal currents, which resulted in intelligible coding of the speech information. However, the presentation of pulsatile stimuli at neighboring channels led to the electrical interaction similar to SAS. Only the implementation of the continuous interleaved sampling (CIS) strategy in the early nineties allowed cochlear implant users to accelerate performance for the first time. Today, some CI users communicating over the phone. Present speech processing strategies that support high levels speech perception are and CIS [14], "n-of-m" [15-17] and advanced combination encoder (ACE) [18-20], spectral peak (SPEAK) [21, 22].

Speech processors of all available implant systems receive their input from the microphone. Frequency components below 1.2 kHz are attenuated with a pre-emphasis filter (slope of - 6 dB/octave) to boost the relatively weak consonants over the dominating vowels below 1.2 kHz.

The output of the pre-emphasis filter is further processed through a bank of bandpass filters to filter the acoustic signal into different frequency bands. The center frequencies of the bandpass filters correspond to different sites along the cochleae as determined by the location of the electrical contacts along the frequency-place map of the tonotopically organized cochlea: low frequency bands are represented by the apical contacts, whereas high frequency bands by the basal cochlear implant electrode contacts.

In each frequency band, the low frequency envelope of the acoustical signal is extracted and transmitted via the RF. The envelope detection is achieved by a rectifier or Hilbert transform, followed by a low-pass filter (cutoff frequency range of 200 to 400 Hz) [23]. Note that this envelope detection encompasses the fundamental frequency of voiced speech sounds. A logarithmic compression maps the wide dynamic range of natural hearing to the electrically evoked hearing [24-26]. The output of the speech processor is used to modulate a carrier, trains of high-rate (1000 pulses/s/electrode or higher, e.g. 2800 to 5600 Hz for HiRes strategy) interleaved electrical pulses [27]. The interleaved delivery of the pulses on neighboring pulse trains is necessary to avoid interaction for overlapping current fields during stimulation. To sum up, the coding strategies extract mainly the frequency range that covers the envelope of speech (<50 Hz) as well as the fundamental frequency and its harmonics up to 200 to 400 Hz. This frequency range also corresponds well with the perception of pitch in CI users, which is up to 300 Hz [28-31].

The n-of-m and ACE strategies are similar in terms of design and performance [16, 17, 20]. They also share similar bandpass filtering, envelope detection and the interlacing stimulation as in CIS. The differentiator between n-of-m and ACE are the "peak picking" scheme: the envelope in

all channels (m) are scanned rapidly prior to each stimulation frame and the number of channels (n) with the highest envelopes are selected for stimulation.

SPEAK is adapted from the n-of-m strategy with similar concepts of bandpass filtering and envelope detection. Its differentiator is that the number of the selected frequency band (n) can vary for different stimuli. SPEAK filters separates the recorded signal into 20 frequency bands, and selects the ones with envelopes exceeding a preset "noise threshold" for each scan. The number of n can vary from 1 to 10, and typically n is around 6.

1.3 Aims and significance

1.3.1 Aim 1: To develop a light delivery system for INS in the cochleae

While CIs restore speech recognition well in quiet listening environments, all users are challenged in noisy listening environments, for tonal languages and for the perception of music [2]. It has been argued that one of the underlying reasons for poor patient performance in challenging listening environments is the poor spatial selectivity of electrical stimulation, which limits the number of independent frequency channels. The spatial specificity relies on a few factors, such as the number and distributions of the ganglion cells, the proximity of the electrode to the neurons. Efforts have been made to increase the number of independent channels for stimulation, including electrode placement close to the inner wall of scala tympani [32-34], multipolar stimulation [4, 35-37] and current steering [38-40]. Current steering refers to the simultaneous discharge of neighboring electrodes to "steer" the current to selected neuron populations,

introducing an additional virtual channel [39-43]. However, this technique does not introduce more independent channels for parallel stimulation [39, 41-44]. In addition to the manipulation of electrical stimulation, a more recent approach for spatially selective neural stimulation is the use of photons [44-46].

1.3.1.1 Photonic neural stimulation

Photons have been used for neurostimulation and neuromodulation in the cochlea. The following approaches and mechanisms have been explored: (1) direct stimulation of the target structure with infrared radiation (infrared neural stimulation or INS), (2) activation of ion channels expressed in neurons with visible light (optogenetics), (3) activation of temperature sensitive ion channels expressed in neurons with heat (thermogenetics), (4) and in the cochlea activation of neurons through mechanical events created by laser radiation. To determine the mechanism involved in photonic stimulation, the interactions between photons and the tissue need to be studied. For the properties of photons, the wavelength, radiant energy, and the temporal properties (duration, pulse shape, repetition rate, etc.) should be considered. As for the tissue being irradiated, the photon absorption, reflection, and scattering should be examined. A detailed summary of the history of optical stimulation research can be found in Richter et al., 2014, Table 1.

1.3.1.2 Infrared Neural Stimulation (INS)

The first successful experiments with a pulsed infrared laser to stimulate nerves has been reported for the sciatic nerve in rats [47, 48]. Wells and coworkers studied diligently the light tissue interactions by using the free-electron laser and identified several suitable wavelengths for safe neural stimulation in the near infrared and infrared [49]. A wide range of wavelengths from ~1,500 to 10,000 nm were explored to find the suitable wavelength, and the wavelength from ~1,860 to 2,100 nm was studied due to the laser availability (holmium:yttrium-aluminum-garnet (Ho: YAG) laser and the Aculight laser). Upon the absorption of the photon by the water, its energy is converted into heat [50] which then depolarizes the cell by changing the membrane capacitance [51, 52]. It has been suggested that this capacitance change was caused by changes in membrane thickness [53] or from small-diameter nanopores in the membrane [54]. It has also been demonstrated that transient receptor potential cation channels of the vanilloid group (TRPV) are involved in INS [55-58]. TRPV are temperature sensitive and are highly calcium selective [59-68]. Previous studies have shown that intracellular calcium homeostasis changes during INS [68-72]. Spatially and temporally confined heating during INS also results in stress relaxation waves [73]. There is an ongoing debate is whether or not the resulting pressure is the dominating effect in cochlear INS. Results have been presented where cochlear INS did not evoke responses in deaf animals [74, 75]. These findings differ from reports that showed responses in deaf animals missing hair cells [76, 77]. The argument for direct interaction between the radiation and the neurons comes from single unit recordings done in the inferior colliculus [78] and masking experiments in the guinea pig [79]. Furthermore, INS evokes ABRs in congenitally deaf mice [80]. One of the mouse models lacks the vesicular glutamate transporter-3 (VGLUT3$^{-/-}$) and lacks glutamate release at the

inner hair cell afferent synapse [79, 81-83]. Different gene-manipulated mice are the *Atoh1$^{f/kiNeurog1}$*

mice [79, 84], which show no ABR response to acoustical stimuli but responses to INS [85].

1.3.1.3 Aim 1: To develop a light delivery system

To achieve INS in cochlear implants, photons must be delivered to selected sites along the

cochlea through a light delivery system (LDS). This can be achieved by inserting into scala

tympani: (1) an array of small optical sources, such as side emitting laser diodes (SELDs) or

vertical cavity surface emitting lasers (VCSELs), (2) a bundle of glass fibers, or (3) a bundle of

polyimide waveguides (Table 1-1). Low H_2O containing glass fibers are too stiff and will damage

the cochlear soft tissue structures during insertion into scala tympani and cannot be inserted at

sufficient lengths [86]. Optical sources and plastic (e.g. polyimide) waveguides remain as a

possible alternative. The current sizes of VCSELs or SELDs, which are powerful enough to

stimulate the neurons, are about 400 µm x 350 µm x 200 µm. For the study presented in Chapter

2, VCSELs (Vixar Inc., Plymouth, MN) were used as laser sources. VCSELs with 1850 nm

wavelength were used to build single-channel and multi-channel optical arrays. The measurements

of VCSELs were small enough to be inserted into the scala tympani of the basal cochlear turn of

guinea pigs for in vivo functional tests.

Table 1- 1 A comparison across three types of LDS: small optical sources, optical fibers, and waveguides.

LDS	Spatial selectivity: 50 individual channels	Fully implantable	Safety requirement

Small optical sources	Can achieve 50 individual radiation sites	Fully implantable: Eliminates the transcutaneous link	Can be encapsulated in biocompatible materials
Optical fibers	Difficult to achieve 50 individual radiation sites (200 μm for each radiation site)	Not fully implantable w/o light sources	Too stiff: break during insertion and damage the cochlea
Waveguides	Difficult to achieve 50 individual radiation sites (200 μm for each radiation site)	Not fully implantable w/o light sources	Biocompatible

To date, prototypes of optical arrays have been assembled with promising measurements for INS in the cochlea. However, no functional test has been reported regarding whether these optical arrays can generate auditory responses. In Chapter 2 of this dissertation, we tested the function of VCSELs arrays in an acute experiment in normal hearing guinea pigs. Auditory responses could be observed with the activation of VCSELs arrays.

1.3.2 Aim 2: Examine the function of speech temporal periodicity and its neural processing

To accommodate the low rate of optical stimulation, a low-rate coding strategy was developed in our lab. The coding strategy is referred as frequency modulated phase coding (FMPC), which could elicit speech perception in CI users (working paper from the Richter lab). A closer look at the pulse rate generated by FMPC showed synchronization to speech temporal cues lower

than 500 Hz, including fundamental frequency, F0, and its harmonics (Figure 1-1). Recent modeling studies have demonstrated that recreating stimulus spectra in the AN rate profiles will not elicit appropriate responses from central neurons, whereas restitution of the F0-synchronized rate profiles may [87-90]. In this thesis, we examine the functional role of the temporal cues within 500 Hz for speech perception and its neural representation. This dissertation demonstrated that the temporal cues within 500 Hz are essential for speech perception in normal hearing listeners under both quiet and noisy listening environment in Chapter 3. The dissertation also uncovered the neural processing of the periodic cues in Chapter 4.

Figure 1- 1. An example of speech sentence processed by FMPC.
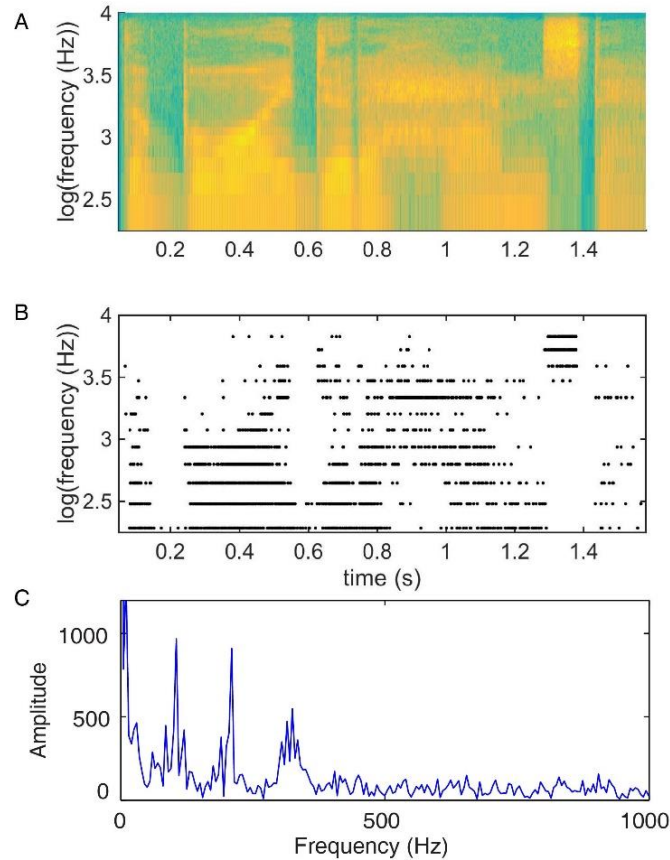
Figure 1-1. An example of a speech sentence processed by FMPC. A shows a spectrum of the

sentence "The boy did a handstand". B shows the electrical pulses generated by the novel algorithm.

Each line represents one frequency band. C shows the spectrum of a selected pulse train at 676 Hz

frequency band that was generated by the vowel "o".

1.3.2.1 Nomenclature for temporal cues of speech

The functional role of the temporal cues has been extensively studied by different research

groups. However, the definition of the temporal cues has not been consistent across studies. The

terms used in the literature to describe the temporal cues of speech include the envelope, temporal fine structure, or temporal periodicity, etc. They describe frequency ranges of magnitude fluctuations of the filtered signal. Rosen (1992) defined the temporal cues based on the frequency ranges as following: envelope (E or TE): the low frequency magnitude fluctuation <50 Hz; periodicity: The 50 to 500 Hz magnitude fluctuation; temporal fine structure (TFS): The high frequency ranges from 600 to 10,000 Hz. Here is an example demonstrating the three ranges of frequencies. A 200 ms section of the natural speech was selected as the example contain a vowel "o". The time waveform and frequency components of this speech signal are shown (Figure 1-2 a, b). The speech signal was filtered into 16 frequency bands up to 10 kHz. The frequency bands centered at 689 Hz was selected. The time waveform and frequency components were also shown (Figure 1-2 c, d). The range of E, periodicity, and TFS were marked (Figure 1-2d).

Figure 1- 2. The temporal cues of speech.
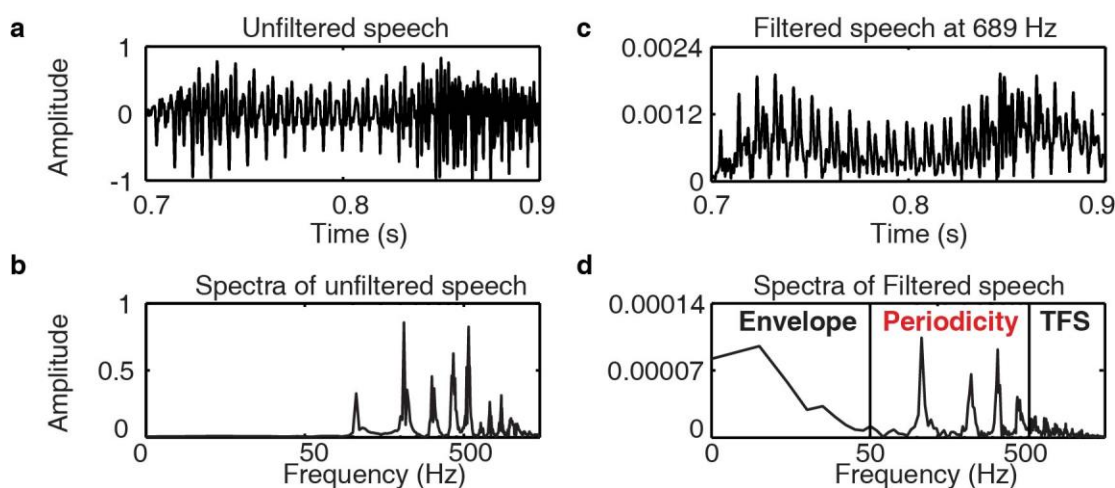
Figure 1-2. A voiced syllable "o" from "the silly boy is hiding" was selected (0.714 to 0.914 s post the onset of each sentence). a and c showed the time waveform of unfiltered speech

and band-pass filtered speech at the center frequencies of 689 Hz. d and d showed the spectra of the unfiltered speech and band-pass filtered speech. The range of envelope, periodicity, and part of the TFS were listed.

"Periodicity" is consistently used to describe 50 to 500 Hz frequency range specifically [91-96]. However, we have recognized that recent literature also refers this range of frequency as part of the TFS. Additionally, terminologies like "envelope periodicity" are sometimes used, as a general descriptive term for the waveform with no explicitly defined frequency range. For example, Hall (1979) used "envelope periodicity" and "periodicity" interchangeably in his article. Due to the inconsistent use of terminology in literature, it is hard to find a universal, accurate term to describe the frequency range 50 to 500 Hz. In order to serve the general readers, we have adopted the term "periodic cues of the TFS". This is because we recognize that recent studies have only addressed the E and TFS when describing temporal cues, but neglect the controversial frequency range from 50 to 500 Hz.

Although E and TFS have been clearly defined as mentioned above and are commonly used in the literature, E and TFS were extracted differently in various studies resulting in inconsistency of their exact frequency ranges. An example of the incoherent use of the terms is shown for two heavily cited studies: Shannon et al. 1995 and Lorenzi et al., 2006. A detailed review of the methods used to extract E and TFS and the corresponding definitions showed substantial differences in the frequency range for the same terminology among studies from different laboratories. We have outlined the differences in detail in Table 1-2, where the difference in the

frequency range for the term "envelope" can be seen. Shannon used a low pass filter with (1) cutoff frequency of 16, 50, 160 or 500 Hz; (2) a roll off slope of -6 dB/octave to extract E. Lorenzi used a Hilbert transform followed by a low pass filter with (1) cutoff frequency of 64 Hz, (2) a roll off slope of -72 dB/octave. Given the sound level used by Shannon and Lorenzi were both 75 dBA, the E and TFS contained different energy at certain frequencies due to the different roll off parameters. For the purpose of calculation, we picked one frequency band that covers 0 to 1024 Hz, similarly to the lowest frequency band in Shonnon (1995) when the number of frequency band was four. For example, Shannon's E (extracted with the low pass filter that has a cutoff frequency at 16 Hz) would contain substantial energy at 128 Hz (57 dBA = 75 dBA - 6 dB/oct * 3 oct, 128 Hz is 3 octaves from 16 Hz), whereas Lorenzi's E contained minimal energy at 128 Hz (3 dBA = 75 dBA - 72 dB/oct * 1 oct). A detailed comparison of sound level at different frequencies of E can be seen in Table 1-2. Of note, the initial Hilbert transform was not considered in the table for Lorenzi's analysis. The sound level should be even lower if Hilbert is considered since Hilbert transform extracts the relatively low frequency components within each frequency band. Similar discrepancies could be seen in other studies.

Table 1- 2 Sound levels (dBA) of "envelope" at different frequencies.

	Lowpass filter parameters	The frequency of "envelope						
		16 Hz	32 Hz	64 Hz	128 Hz	256 Hz	512 Hz	1024 Hz
Lorenzi	Cutoff at 64 Hz Rolloff -72 dB/oct	75	75	75	3	0	0	0

Shannon	Cutoff at 16 Hz Rolloff -6 dB/oct	75	69	63	57	51	45	39
	Cutoff at 50 Hz Rolloff -6 dB/oct	75	75	75 - 69	69 - 63	63 - 57	57 - 51	51 - 45
	Cutoff at 160 Hz Rolloff -6 dB/oct)	75	75	75	57	75 - 69	75 - 69	63 - 57
	Cutoff at 500 Hz Rolloff -6 dB/oct	75	75	75	75	75	75 - 69	63 - 57

Table 1- 3 The estimated frequency range of "Envelope" for a few existing studies.

Temporal cues defined by Rosen (1992)	Envelope <50 Hz	Periodicity 50 – 500 Hz	Temporal fine structure > 500 Hz
Shannon et al., 1995: low pass cut off at 16, 50, 160, 500 HZ; 3 dB/octave roll off (4 frequency bands)			

Turner et al., 1995, 1999: low pass cut off at 30 HZ; roll off not described	▪▪▪▪▪▪▶		
Loizou et al., 1999: low pass cut off at 400 HZ; roll off not described	▪▪▪▪▪▪▪▪▪▪▪▪▶		
Souza et al., 2006: low pass cut off at 50 HZ; roll off not described	▪▪▪▪▪▶		
Lorenzi et al., 2006: low pass cut off at 64 Hz; 75 dB/octave roll off (16 frequency bands)	⟶		

Note: The dash line was used due to unknown roll off of the low pass filter.

Moreover, Table 1-2 and 1-3 shows, for example, the sound level of E within one frequency band. Additional discrepancies in sound level can also be expected from the output from different selections of auditory filters. Assuming that E and TFS added together equals the output of each frequency band, TFS must be different regarding its frequency range. Other studies adopted different types of filters and different methods to extract E and TFS. For example, Smith (2002) applied a Hilbert transform on a bandpass filtered signal with a bandwidth of 45 Hz, while Dubbelboer, F. & Houtgast (2007) used a Gaussian-shaped wavelet analysis on a bandpass filtered signal with a bandwidth of ¼ octave. It is important to understand the precise frequency range of given E and TFS to accurately interpret and compare the results across studies.

Aim 2a. To examine whether the temporal periodic cues are essential for speech perception in both quiet and noisy environment

Direct and fundamental neural correlates to the temporal cues have been explored in animals by measuring the discharge patterns of auditory neurons, in particular, the timing of neural discharges or the phase locking of the neurons [86]. Phase locking is a phenomenon that is preserved along the auditory pathway from the cochlea to the cortex [97-105]. It is referred to as the tendency for auditory neurons to fire within a well-defined time window relative to a period of the stimulating frequency. This is particularly evident in the neural response to low-frequency pure tones [86]. Phase locking pattern is also seen in auditory nerve fibers in response to periodic sounds, such as modulated frequencies and steady-state vowels. Importantly, the TFS of natural speech sentences contains multiple representative frequencies, including the fundamental frequency (F_0)

and its harmonics. Auditory neurons may discharge at the rate of these frequencies [97, 106-108], i.e., phase lock to these frequencies. Thus, the phase locking pattern could serve as a tool in examining the neural coding of the TFS in speech.

While phase locking could be used to investigate the neural components of TFS, reconstructed speech sounds with distorted temporal periodicity could be used to investigate its perceptual functions. Previous studies have frequently manipulated speech sounds to assess the perceptual role of different acoustic cues. The modified speech could be achieved by keeping the acoustic cues of interest while disrupting or replacing other acoustic components, and vice versa. For example, studies examining E have reconstructed acoustic signals by replacing spectral cues with modulated sinusoids or pulse trains [109] or spectrally matched noise [110]. The E is usually extracted by a Hilbert transformation or by the low pass filtering of the rectified bandpass filtered speech signal. Through dividing the E from the original signal or peak clipping, TFS cues could be isolated. An alternative method to distort speech is based on the short-time Fourier transform (STFT) [111] or Wavelet transformation [112]. According to Kazama, Short-Time Fourier Transform (STFT) is the ideal tool to engineer target distortions at any selected frequency range by using different time windows (Kazama et al., 2010). Thus, STFT was used to systematically distort the frequency range from 50 to 500 Hz. We have discussed this in the method section of Chapter 3. Additionally, six frequency bands out of 64 were used to mimic the n-of-m strategy of cochlear implant speech processors. Cochlear implant users are reported to have only four to eight independent frequency bands. Shannon used four frequency bands, but the E contained energy at high frequencies from 50 to 500 Hz, and even up to 1000 Hz. The frequency range for the envelope selected by Lorenzi contained little energy from 50 to 500 Hz or higher, but 16 frequency bands

were used. Our approach contained E below 50 Hz at six frequency bands, which best represents the input to cochlear implant users.

In Chapter 3 of this dissertation, we focused on the neural mechanism of TFS processing by combining neural recordings from individual auditory neurons in guinea pigs with psychophysical measures in human subjects. The TFS cues of sentences from the Hearing in Noise Test (HINT) and the QuickSIN™ test were modified to various extents. The modified speech signals were used for speech perception tests in normal hearing listeners in both quiet and noisy conditions. These same sentences were used to evoke neural activity in well-identified auditory neurons of the inferior colliculus (ICC) in guinea pigs. ICC was selected as the target structure due to its robust representation to the frequency range of the periodicity cues. Through this direct comparison between behavioral and neural data, our results bridged the perceptual role and neural coding of TFS cues in speech, allowing us to uncover the essential neural mechanism for TFS perception.

Aim 2b. To examine the neural processing of the periodic cues of the TFS in the midbrain

Harmonic signals are commonly present in human speech, music, and animal vocalizations [91, 113]. They are characterized by the combination of a fundamental frequency (F0) and their integer multiples and are critical for speech perception [114, 115]. The amplitude and phase of these harmonic components contribute to the temporal features of the signal, which is always

periodic at the frequency of F0 even when F0 is not present in the signal. Harmonic signals can evoke strong pitch sensation at F0. The neural coding of pitch sensation has been extensively studied at the level of individual auditory neurons in the cochlea [98, 99, 116-119], cochlear nucleus [120-123], inferior colliculus [124, 125] and the cortex[126]. It has been suggested that firing rate and cochlear tonotopic organization, i.e. the rate-place coding, is not sufficient to code harmonic signals. For example, in response to low-frequency resolved harmonic signals, cochlear response peaks at these harmonics based on its tonotopic organizations. However, such rate-place coding cannot encode high-frequency unresolved harmonics. The neural coding of harmonic sounds also requires temporal coding, i.e. the neural synchronization or interspike interval distributions or phase locking [127-129]. Phase locking refers to the timing of neural discharges occurring at a certain phase of cyclical waveforms in response to periodic sounds. This temporal coding, which can process both resolved and unresolved harmonic signals. Both rate-placing coding and temporal coding were observed in the cochlea [98, 99, 116-119], and the cochlear nucleus [120-123]. However, the predominant temporal coding in the auditory periphery changes to the rate coding in the cortex [126, 130]. It has been suggested that the majority of the units in cortex encode only the lower frequency range relevant for periodicity pitch [131, 132]. It has also been suggested that cortical units processing slow modulations with rate-coding, which are restricted to a low-frequency cortical region near the anterolateral border of the primary auditory cortex [126]. The inferior colliculus (ICC), which receives input from several peripheral auditory nuclei and relays information to the auditory cortex, is likely the stage where the transition occurred [133].

The neural temporal coding of harmonic signals has been widely examined with amplitude modulated sounds [113]. Phase-locked responses to low frequency modulation have been recorded in neurons located at the cochlea [98, 99, 116-119], cochlear nucleus [120-123], inferior colliculus [100], medial geniculate body [134, 135], and at various areas of the auditory cortex [136, 137]. Auditory nerve fibers show robust phase locking to low frequency tones up to several kilohertz (eg. 9 kHz in barn owls [138], 4-5 kHz for cats [118], 3.5 kHz in guinea pigs [99, 139]). And the upper limit of phase locking is species dependent [140]. Although the relationship between phase locking and BF has not been explored in great detail, existing evidence suggests great effects of BF on phase locking [141, 142]. The robust phase locking degrades in the level of ICC: 60-70% ICC neurons show phase locked responses to pure tones and the cut off is around 1 kHz in guinea pigs [100]. Similar degradation in synchronization at ICC has been observed in other species [113]. Phase locking to modulations was seen in 30% of ICC neurons in anurans [143]. Langner and Schreiner's data in cats showed that ~ 75% of the single units in ICC were tuned to a best modulation frequency when measured by their firing rate, in contrast to ~33% of the single units that were tuned to a best modulation frequency in their synchronization [124]. More recently, Su and Delgutte showed non-tonotopic rate coding in the ICC using harmonic complex sounds [133]. These results indicate the temporal coding is partially transformed into rate-coding at the level of ICC.

The neural temporal coding of harmonic signals has also been studied with speech tokens, such as synthesized vowels [99, 114, 139, 144, 145]. Auditory nerve fibers were suggested to synchronize to the F0 and certain high-frequency formants up to several kilohertz, as well as certain distortions [97, 101-105, 146, 147]. Phase locking to these frequency components is related

to the amplitude of the harmonics and the neurons' best frequencies (BFs). At higher auditory centers, such as the ICC, the neural synchrony is restricted to the low frequency harmonics of vowels. In response to a steady-state approximation to the vowel /a/, most of the low frequency neurons in cats at the ICC (BFs<4 kHz) are strongly phase locked to the F0 and its low harmonics (<1 kHz), but not to the higher formants [144]. This confirms the significant loss of temporal coding at the midbrain. Only ~47% of the guinea pig ICC neurons (BFs<8 kHz) synchronize to F0, and only ~5% ICC neurons phase lock to the time-varying F0 in tonal language syllables (Mandarin syllable /ba/ with four lexical tones: flat, rising, falling then rising and falling) [148]. It remains unknown how many ICC neurons phase lock to F0 in response to natural English speech sentences, whose F0 is non-steady but less variable than tonal language. Our study systematically examined the phase locking pattern in ICC neurons over a wide range of BFs with natural English speech in guinea pigs. We also examined the neural representation of speech E through neural firing rate, particularly identifying how each neuron synchronizes with a given integration period and range of frequencies.

In Chapter 4 of this dissertation, we adopted natural speech sentences as acoustic stimuli and recorded action potentials from individual ICC neurons. We systematically examined the phase locking pattern in ICC neurons over a wide range of BFs with a natural English speech in guinea pigs. We also examined the neural representation of speech E through neural firing rate, particularly identifying how each neuron synchronizes with a given integration period and range of frequencies.

Chapter 2

Multichannel optrodes for photonic stimulation

Yingyue Xu[1,2*], Nan Xia[3*], Michelle Lim[1], Xiaodong Tan[1], Minh Ha Tran[1], Erin Boulger[1], Fei Peng[4], Hunter Young[1], Christoph Rau[5], Alexander Rack[6], Claus-Peter Richter[1,2,7,8]

[1] Northwestern University Feinberg School of Medicine, Department of Otolaryngology, 320 E. Superior St, Searle 12-561, Chicago, IL 60611, USA.

[2] Department of Communication Sciences and Disorders, Northwestern University, 70 Arts Circle Drive, Evanston, IL 60208, USA.

[3] Institute for Digital Medicine and Computer-assisted Surgery, Qingdao University, Qingdao 266005, China.

[4] Bioengineering College, Chongqing University, Chongqing 400044, China.

[5] Diamond Light Source Ltd, Diamond House, Harwell Science and Innovation Campus, Didcot, Oxfordshire, OX11 0DE, UK

[6] European Synchrotron Radiation Facility, Structure of Materials Group - ID19, CS 40220, 38043 Grenoble Cedex 9, France

[7] Department of Biomedical Engineering, Northwestern University, 2145 Sheridan Road, Tech E310, Evanston, IL 60208, USA.

[8] Hugh Knowles Center for Clinical and Basic Sciences in Hearing, Northwestern University, 70 Arts Circle Drive, Evanston, IL 60208, USA.

*authors equally contributed to the work.

Address:

Claus-Peter Richter, M.D., Ph.D.; Northwestern University Feinberg School of Medicine; Department of Otolaryngology; 320 E. Superior St, Searle 12-561; Chicago, IL 60611-3008; U.S.A.; Phone: (312) 503 1603; FAX: (312) 503 1616; e-mail: cri529@northwestern.edu.

2.1 Abstract

An emerging method in the field of neural stimulation is the use of photons to activate neurons. The possible advantage of optical stimulation over electrical is attributable to its spatially selective activation of small neuron populations, which is promising in generating superior spatial resolution in neural interfaces. Two principal methods are explored for cochlear prostheses: direct stimulation of nerves with infrared light and optogenetics. This manuscript discusses basic requirements for developing a light delivery system (LDS) for the cochlea and provides examples for building such devices. The proposed device relies on small optical sources, which are assembled in an array to be inserted into the cochlea. The mechanical properties, the biocompatibility, and the efficacy of optrodes have been tested in animal models. The force required to insert optrodes into a model of the human scala tympani was comparable to insertion forces obtained for contemporary cochlear implant electrodes. Side emitting diodes are powerful enough to evoke auditory responses in guinea pigs. Chronic implantation of the LDS did not elevate auditory brainstem responses over 26 weeks.

2.2 Introduction

2.2.1 Cochlear implants and their challenges

Cochlear implants (CIs) are considered one of the most successful neural prostheses. Today about 350,000 individuals with severe-to-profound hearing loss have received a CI to restore some of their hearing. However, the performance of individual users varies largely. While some patients are able to communicate over the phone in different languages, others receive little benefit from CIs. For all CI users, noisy listening environments and music perception constitute a challenge [1, 2]. It has been argued that performance could be improved by reducing the interaction between neighboring CI electrode contacts, subsequently creating more independent channels for stimulation. More spatially selective stimulation with electric current can be achieved through multipolar stimulation, where multiple electrode contacts are used to narrow the current field [35, 37, 149]. Another approach to increasing the number of different pitch percepts is called current steering [38-40]. In this approach, neighboring electrodes are used simultaneously to "steer" the current to selected neuron populations between the two contacts. However, this technique does not introduce more independent channels for parallel stimulation [39, 40, 42-44].

2.2.2 Optical stimulation

More recently, the use of photons has been suggested as a novel approach to evoke responses from small populations of neurons [44-46] because optical radiation can be delivered spatially selectively [49]. It is anticipated that optical stimulation decreases interactions between neighboring channels allowing for the development of neural prostheses with enhanced neural fidelity. Two methods for direct neural stimulation with light are currently considered:

optogenetics and infrared neural stimulation (INS) [52, 150, 151]. Optogenetics requires the delivery of a viral vector to express photosensitive ion channels in the membrane of the spiral ganglion neurons (SGNs) [152, 153]. INS does not require such treatment because during INS, the fluid in the target tissue absorbs the photons and the energy is converted into heat [55, 154-156]. Spatially and temporally confined heating evokes action potentials in the SGNs (see below for mechanism). While both methods appear promising, they also have challenges. For optogenetics, the neurons must be manipulated genetically. This requires targeting of a selected population of neurons with a viral vector to induce stable expression of light-sensitive ion channels. The rate by which the ion channel or optogenetic tool is expressed is crucial since low expression of the optogenetic tool will require larger photon flux rates and high expression of the ion channel may damage the cell. Moreover, tissue in the beam path largely scatters the incident photons resulting in broad response profiles and a significant reduction of the transmitted radiant energy. Published results have shown that the energy required to evoke an action potential on the murine auditory nerve is about 10-times larger than for electrical stimulation [157]. The challenge for INS is the delivery of heat to the target structure(s), which needs to be removed or dissipated to prevent thermal damage during stimulation. Tissue heating limits the rate of stimulation to about 250 pulses per second (pps) at a maximal radiant energy of 25 μJ/pulse [158-160]. Optical sources, small enough to be inserted into the cochlea, have a low wall-plug efficiency (ratio in converting electrical power to optical radiation) [46]. The energy required to stimulate with INS is about 100-times larger when compared to electrical stimulation [46, 161]. Optoacoustic events resulting from INS must be considered for patients with residual hearing.

When comparing optogenetics and INS, the wavelengths of the radiation used for each of the two methods should also be considered. The incident radiant energy is reduced by the tissue in the beam path through scattering and absorption of the photons. At wavelengths used for optogenetics, $\lambda<1064$ nm, the extinction coefficient for the radiation is governed by the scattering of the photons, whereas above $\lambda=1064$ nm, the absorption becomes the dominant factor in tissue. Hence, for optogenetics, tissue reduces the transmitted radiant energy and broadens the beam by scattering. For INS, the energy is mostly reduced through absorption of the photons and less through scattering [162, 163].

2.2.3 INS - neurons are activated by temporally and spatially confined heating

One of the first reports on laser irradiation as a method to stimulate neurons came from Fork's study on *Aplysia Californica* [163]. Irradiation of the tissue with blue ($\lambda=488$ nm) light (spot size = 10 µm) evoked action potentials at stimulus levels above 12.5 mW radiation power [47]. Wells and coworkers studied light tissue interactions by using the free-electron laser in great detail. They determined radiation wavelengths that could be used for safe neural stimulation, which are in the near infrared and infrared [47]. One of the wavelength ranges for which optical sources exist for stimulation is between 1840 and 2100 nm. Upon the absorption of the photons by the water, their energy is converted into heat [164], which then evokes an action potential. Temporally and spatially confined heating depolarizes the cell by changing the membrane capacitance [165-168] resulting in a depolarizing inward current. The change in capacitance might result from changes in membrane thickness [169] or from small-diameter nanopores in the membrane [54]. Furthermore, it has been demonstrated that transient receptor potential cation channels of the

vanilloid group (TRPV) are involved [54, 56-58]. They are temperature sensitive and are highly calcium selective [59-68]. Published results demonstrated that intracellular calcium homeostasis changes during INS [68-72, 170]. Spatially and temporally confined heating, which occurs during INS, also results in stress relaxation waves [73]. Those pressure waves could vibrate the basilar membrane and evoke auditory responses through stimulation of remaining inner hair cells. Results have been presented where cochlear INS did not evoke responses in deaf animals [74, 75]. The authors concluded that INS in the cochlea only originates from the generation of a pressure wave. Those findings differ from reports that showed responses in deaf animals missing hair cells [76, 79] and in congenitally deaf mice [80]. One of the deaf mouse models lacks the vesicular glutamate transporter-3 (VGLUT3$^{-/-}$) and does not release glutamate at the inner hair cell afferent synapse [79, 81, 82]. This mouse model shows no auditory response to sound stimuli, but responds to INS, indicating direct interactions between INS and SGNs. Other gene manipulated mice which show no ABR response to acoustical stimuli, but respond to INS [83] are the *Atoh1-cre; Atoh1$^{f/kiNeurog1}$* mice [84, 85].

To use INS in a CI, photons must be delivered to selected sites along the cochlea. This can be achieved by inserting light delivery systems (LDSs) into scala tympani of the basal turn of a cochlea. LDSs can be arrays of optical sources, such as side emitting laser diodes (SELDs) or vertical cavity surface emitting lasers (VCSELs), bundles of glass fibers, or bundles of polyimide waveguides. Low H_2O containing glass fibers are not considered as LDS because they are too stiff and break easily if they are larger than 50 µm in diameter. Stiff optical fiber bundles will also damage the cochlear soft tissue structures during insertion into scala tympani and cannot be inserted at sufficient lengths [171]. Polyimide waveguides are flexible and biocompatible.

However, challenges remain in coupling the light sources and waveguides. Moreover, the size of the waveguides limits the number of individual optical stimulation sites along the spiral ganglion [78].

This paper discusses how to build and test optrodes with small optical sources, which are powerful enough for INS.

2.3 Methods

2.3.1 Light sources

Currently, three different types of light sources have been used to fabricate optrodes, vertical cavity surface emitting lasers (VCSELs), side emitting laser diodes (SELDs) and micro-light-emitting diodes (μ-LEDs). The light sources are available in sizes that can be assembled into arrays suitable for insertion into scala tympani (Figure 2-1). Considering the frequency place map in the human cochlea (organ of Corti and spiral ganglion) [172, 173], 50 light sources could be assembled into a 24 mm long array and provide a frequency resolution of about 1/8 of an octave [79, 80]. Our most recent arrays feature 15 light sources on a 24 mm long array. The number of optical sources will be increased in future optrodes by placing the optical dies closer together.

Figure 2- 1. Different types of light sources.
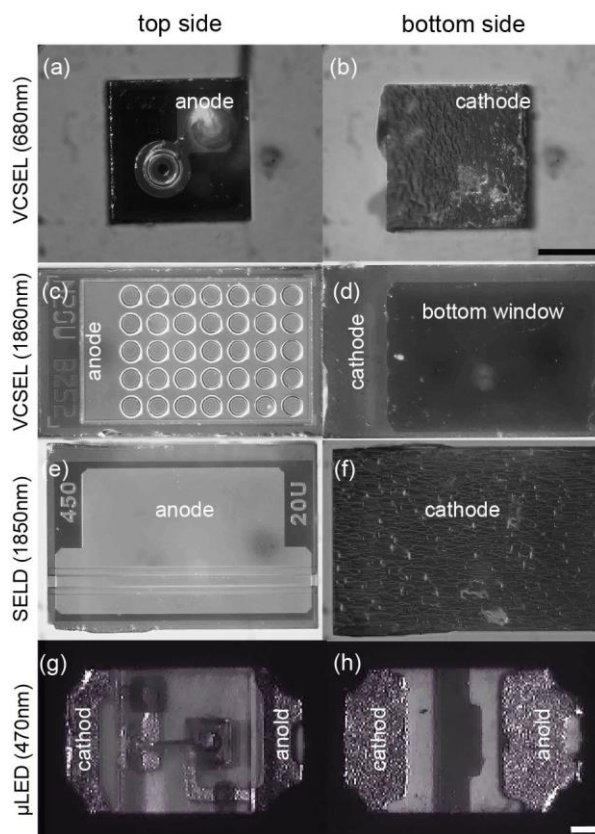
Figure 2-1. (a) shows the top and (b) the reverse side of a VCSEL (λ=680 nm). The dimension is 250 x 250 x 200 µm3 and its maximum output power is about 4 mW. (c) and (d) show the top and bottom of a 5 x 7 VCSEL array, (λ=1860 nm). The dimension is 450 x 250 x 200 µm3 and its maximum output power is about 7.5 mW. Each circle in (c) represents one VCSEL. (d) shows the light emitting window and the cathode; (e) and (f) show a SELD (λ=1850 nm). The dimension is 450 x 350 x 100 µm3 and its maximum output power is about 50 mW. (g) shows the original appearance of a blue µLED (λ=470 nm) before being resized. The dimension is 1000 x 600 x 200 µm3 and its maximum output power is about 34 mW. Scale bars are the same for (a)-(f) (shown in (b)) and (g)-(h) (shown in (h)): 100 µm. The power ratings are given for continuous wavemode operation.

2.3.2 Optrodes and hybrids

The first step in the fabrication of the multichannel optrode was connecting the cathode of the light sources with conductive silver epoxy (EPO-TEK H20E, Epoxy Technology Inc., Billerica, MA) to a 125 μm diameter silver wire (Figure 2-2 (a), (c), (e)). The distance between the sources could be adjusted as needed. The silver wire, which could be replaced by strips of platinum or silver foil, also serves as a heat sink. In the second step, a 25 μm diameter Teflon coated platinum wire was connected to the anode of each source using conductive silver epoxy. Note, wire-bonding has been tested as an alternate contacting method (Figure 2-1(a)), however, the connection is fragile and using epoxy resulted in more reliable connections of the light sources. Following the assembly of the optrode, the function of each light source was tested before and after the optrode was embedded into silicone. For the silicone embedding, the optrode was placed in a custom fabricated mold. The mold was filled with Silastic (MDX4-4210, Medical Grade Elastomere, base and curing agent (LOT 0006932899, Dow Corning Corp., USA)) and was allowed to cure overnight in an oven at 60°C. After the silicone was solidified, the electrode was removed from the mold (Figure 2-2. (b), (d), (f)) and the wires of the optrode were extended to about 10 cm. Then, the optrode could be inserted into a Tygon® Micromix flexible microbore plastic tubing with an inner diameter of 1 mm and an outer diameter of 1.8 mm (LOT 507206, Saint-Gobain Performance Plastics, Portage, WI) and was connected to a transcutaneous connector (Figure 2-2(b)). During implantation, the transcutaneous connector was secured with Ethilon 3.0 (Ethicon, Cincinnati, OH) to the skin incision (Figure 2-2(b)).

Figure 2- 2. The figure shows optrodes fabricated with small optical sources.
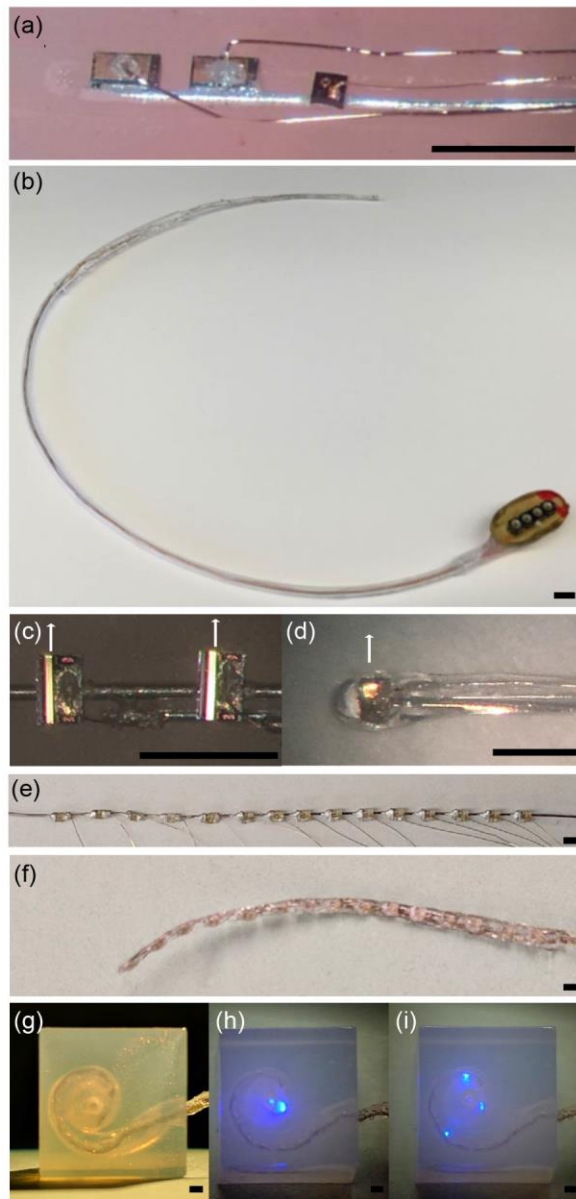
Figure 2-2. Panel (a) is a three-channel optrode. The two infrared sources in (a) are VCSELs (λ=1860 nm, larger dies to the left). A red source (λ=680 nm, smaller die to the right) serves as a pilot light, which helps to orient the implant. Panel (b) shows the completed electrode ready for implantation. At the bottom right, the transcutaneous connector is shown. Panel (c) is a picture of a two-channel optical array made of SELDs. The size of the given emitters is 450 µm x

300 μm x 100 μm. The anode of each emitter was connected with a thin gold wire, and then to a Teflon-insulated silver wire. The cathodes of both emitters were connected to a single silver wire, which also served as a heat sink for the light sources. The direction of the light emission is indicated by the arrow. Panel (d) is a picture of a silicone-embedded single-channel optical array made with a SELD. The anode and cathode connections are the same as shown in panel (c). The array is embedded in silicone. Panel (e) shows an array of 15 μLEDs connected with a silver wire to the cathodes and platinum wires to the anodes. Panel (f) shows an optrode made with 15 light sources. Panel (g) shows the insertion of this optrode into a human scala tympani model. Panel (g) and (i) show the radiation of a single μLED and multiple μLEDs. Scale bars = 900 μm.

A different optrode design used Flexible Printed Circuit Board (FPCB) technology. The single layer FPCB was designed as the light source carrier, which renders the optrode fabrication process much easier. The substrate for the cochlear implant must be soft, flexible, and biocompatible. It has been demonstrated that polyimide polymers are suitable in stiffness (see also discussion), are biocompatible [46, 173-177], and were selected for the support base and the insulation cover layer. Copper was selected as the conductive material. Adhesive films provide the material to bond the copper foil to the base film. The epoxy and copper contacts were further coated with silicone for biocompatibility.

To fabricate the multichannel optrode carrier, a 25 μm-thick copper foil was laminated on the upper surface of the polyimide substrate. Unwanted copper was etched from the copper layer, such that the resulting wire width was 80 μm. To isolate each channel, a 25 μm-thick polyimide film was laminated on the surface. Subsequently, the polyimide film applied for insulation was

etched away on the top of the light source mounting areas and solder joints. Light source mounting areas and solder joints were further improved by electroplating a 25-µm-thick gold layer on the contact areas (Figure 2-3(a)). Dictated by the number of current sources of our portable diode driver system, we only fabricated three channel optrodes. The number of contacts, however, can easily be expanded. Figure 2-3(b) shows the FPCB carrier and its tip. Figure 2-3(c) shows the FPCB-based optrode with a connector. Figure 2-3(d), (e) and (f) show the FPCB carrier and a three-channel optrode made with VCSELs, µLEDs and SELDs. This carrier can also accommodate other light sources and metal contacts for electrical stimulation.

Figure 2- 3. The figure shows the fabrication of the optrode based on FPCB technique.
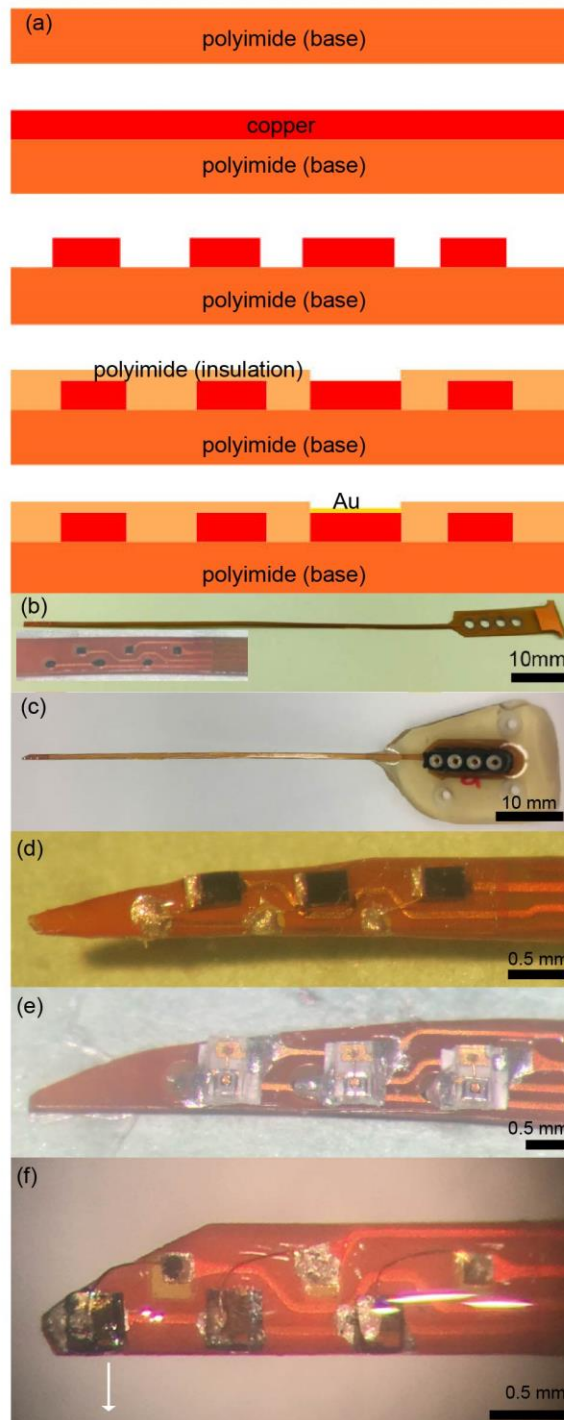
Figrue 2-3. Panel (a) shows the vertical fabrication structure and materials of the light source carrier; (b) shows the FPCB carrier and the tip. The light source mounting area on the tip

is 100 µm × 100 µm; the distance between two channels is 1 mm; the width of the carrier is 0.75 mm and the thickness is 100 µm. (c) shows an FPCB-based optrode with a connector at the end; (d) shows the tip of an FPCB-based optrode with 3 VCSELs. (e) shows the tip of a FPCB-based optrode with 3 µLEDs. (e) shows the tip of a FPCB-based optrode with 3 infrared SELDs.

2.3.3 Insertion force measurements

The insertion force of the electrodes and optrodes was measured in five cadaveric cat and human cochleae, as well as in a matrix printed model of the human scala tympani (~0.016 mm resolution, Objet 260vs Dental, Fisher Unitech Corporation, Chicago, IL). The human cochleae were obtained through the Anatomy Gifts Registry. Each cochlea was accessed via a retro-facial approach and a cochleostomy was created 0.5 to 1 mm from the round window. This allowed the insertion of five short custom fabricated optrodes (with <5 contacts, Table 2-1). Four long arrays (with 10+ contacts, Table 2-1) were tested in the human scala tympani model. For the experiments, the electrodes/optrodes were either mounted on a Narishige electrical step-motor (MM108; Narishige, Japan) or an LTS150 translation stage (Thorlabs, Newton, NJ) to advance the optrode with constant speed at 0.553 mm/s or 0.25 mm/s respectively. At the same time, the insertion force was measured with a Mark-10 Digital Force Gauge (Model M5-012, Mark-10 Corporation, Copiague, NY). The MESUR Lite by Mark-10 software was used to acquire the data at a sampling rate of 10 Hz. The electrode advancement was always monitored through a microscope and was stopped immediately when the electrode would not advance.

2.3.4 Testing in an animal model

All animal procedures were carried out in accordance with the NIH Guide for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee at Northwestern University.

Evaluation in cats

To assess the effects of the optrodes on cochlear function after implantation, auditory brainstem responses (ABRs) to acoustic clicks and pure tone bursts were monitored before surgery and at several time points after surgery and implantation. For the procedure, the cats were sedated with Telazol (5–10 mg/kg, intramuscular) and were given atropine (0.04 mg/kg, subcutaneous). Vitals, such as heart rate, breathing rate, and O_2 saturation, were monitored with a Bionet BM3 vet system (Bionet America, Inc. Tustin, CA USA). Body temperature was maintained with a water-based heating blanket (T/Pump Localized Therapy system, Stryker Global Headquarters, Kalamazoo, MI, USA). Three needle electrodes were placed under the skin to measure ABRs by subtracting ipsilateral mastoid from vertex potentials measured relative to a ground electrode placed in the neck. The contralateral ear was blocked during testing to reduce any possible acoustical crosstalk. Acoustic stimuli were generated by a voltage command presented at a rate of 4 Hz to a Beyer DT770Pro headphone, which was calibrated with a Brüel and Kjær 1/8-inch microphone (Norcross, GA). The speculum of the speaker was placed directly in front of the ear canal (quasi free field). The carrier frequency of the tone bursts started at 32 kHz and was decreased by 2 steps/octave over 5 octaves. The maximum sound level at each frequency varied between 71 dB and 101 dB (re 20µPa), depending on the frequency. For each frequency, the sound level was decreased stepwise by 5 dB until a visible ABR could not be seen to determine ABR thresholds. The ABR electrodes were connected to a differential amplifier (ISO-80, WPI). The

amplifier has a high-input impedance ($10^{12}\,\Omega$) and was set to 80 dB amplification. Further amplification (10 x) and bandpass filtering (0.3 to 3 kHz, -48 dB/octave) of the signal was performed by a digital filter, an IP90 (Frequency Devices, Ottawa, IL). The sampling rate was 250 kHz and 1024 trials were averaged for each stimulation. The threshold was defined as an ABR waveform that was visible above the noise floor of the recordings. After 1024 averages, the noise floor was typically 0.5 µV (peak-to-peak). At the conclusion of the hearing test, each animal recovered from anesthesia was returned to its home cage.

Cochlear implantation in cats

In preparation for the implantation surgery, each animal was premedicated with Telazol (2–4 mg/kg, intramuscular), butorphanol (0.4 mg/kg, subcutaneously), and atropine (0.04 mg/kg, subcutaneously). Intravenous catheters (22G) were placed in the left and right cephalic veins and Ringer's solution containing 2.5% dextrose was given throughout the length of the procedure. Anesthesia was maintained with isoflurane (1%–3%). Only the left ear was implanted with the optrode. The surgical area was aseptically prepared. A ''C'' shaped incision was made behind the left pinna and the bulla was surgically accessed. An opening, approximately 5 x 5 mm^2, was created in the bulla with a motorized drill (Micro-Torque II, WPI) and a 3 mm cutting drill bit to visualize the basal turn of the cochlea. The cochleostomy was then made with a cutting drill bit (1 mm) attached to the motorized drill. The optrode was inserted about 5 mm through the bulla and the cochleostomy into scala tympani of the cochlea (Figure 2-4 (a)). The optrodes were then secured at the bulla with acrylic. The acrylic not only secured the optrode, but also sealed the bulla (Figure 2-4(b), (c)).

A small incision, 1 cm long, was made in the skin between the scapulae. The wire bundles to the optrodes were tunneled under the skin from the bulla to the scapular incision, where the electrical connector was sutured to the skin (Figure 2-4(d)). The incisions were closed in several layers with interrupted sutures. Post-operatively, the animal was monitored daily and received buprenex (0.005–0.01 mg/kg, subcutaneous, 2x/day for 2–3 days) and meloxicam (0.1 mg/kg, oral, 1x/day for 3–4 days) for pain management. No vestibular deficits were seen in any of the animals.

Figure 2- 4. The figure shows the implantation of the optrode into a cat cochlea.
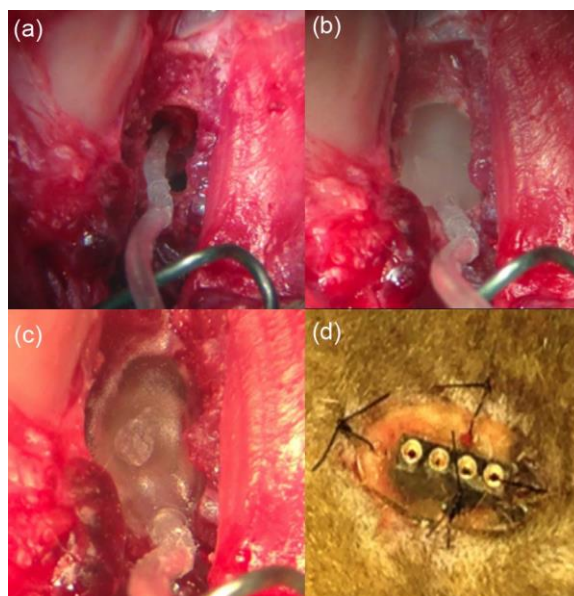
Figure 2-4. (a): The optrode was inserted into a cat cochlea through the cochleostomy. (b): The optrode was fixed to bulla with dental acrylic. (c): the second layer of dental acrylic. (d): The transcutaneous connector was secured onto the lower neck skin.

Acute laser test in cat and guinea pig cochleae

Infrared optrodes were tested acutely in normal hearing cat and guinea pig cochlea. Guinea pig surgery has been described previously. After the cochleostomy was created with a hand-drill, a two-channel infrared optrode with high output power VCSELs or SELDs was inserted into the animal cochlea. While inserting the optrode, care was taken such that the light-emitting window was facing the modiolus. VCSELs had a center wavelength of 1860 nm, SELDs of 1850 nm and were operated at 100 µs pulse duration and 100 pps repetition rate. The test current levels were from 0 to 600 mA. The corresponding voltage ranged from 0 to 2 V. The maximum radiant energy was 20 µJ/pulse (measured above the optrode in air using a Coherent J-50-LP-1A energy sensor). Compound action potentials (CAPs) were recorded with an electrode placed at the round window. The recording system was the same as for the ABR recordings, except that the amplification for the CAP measurements was 60 dB.

2.3.5 Imaging at the European Synchrotron Radiation Facility

Tomographic data sets were acquired at beamline ID19 of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Due to the 150 m-length of the beamline, excellent coherent properties of the X-ray wave fronts at the position of the experiment allow for very sensitive imaging by means of propagation-based phase contrast. In order to reduce dose to the sample tissues while maintaining a high contrast, the beamline's single-harmonic undulator u13 (26.3 keV) was chosen. Besides a 1 mm-thick diamond, a 2.8 mm-thick aluminum absorber and a 0.5 mm-thick polished Beryllium exit window, the beamline was operated optics-free (pink). For detection, an in-house-developed indirect system was used, combining two commercial lenses (Hasselblad) in tandem-design. Using the ratio of the focal distance of the two lenses, here 100

and 200, the magnification was set to 2x. The lenses project the luminescence image of a 500 µm-thick LuAG:Ce (Ce-doped Lu3Al5O12) single-crystal scintillator onto the sensor of a commercial camera (pco.edge, type: 5.5, 2520 x 2160 pixels, each 6.4 µm pixel size). At the conclusion of the experiment, the projections were used to reconstruct the samples. Custom written phase retrieval software was used for the reconstructions [9].

2.3.6 Statistics

Average and standard deviations were calculated. Statistical analysis was completed on the ABR data to determine any significant elevation of threshold following implantation. An ANOVA was used to determine statistical significance, with the null hypothesis indicating no threshold difference between the two conditions. A one-tailed test was used for the post-operative measurement since a threshold decrease following cochlear implantation was highly unlikely.

2.4 Results

2.4.1 In vitro insertion of test electrodes

The five panels of Figure 2-5 illustrate the insertion process of a sham optrode into a cadaveric cat cochlea. The sham optrode has five optical sources but the sources are only connected to the backbone. In Figure 2-5(a) the electrode tip is just inserted through the cochleostomy and in Figure 2-5(d), the entire optrode was inserted. After optimizing the optrode's shape, functional ones were assembled. The longest insertion depth of an optrode in the cat cochlea was 6 mm, which was acceptable for 5 red VCSELs, or 4 infrared VCSELs or SELDs arrays, or 4 blue LEDs.

Figure 2- 5. The figure shows the insertion of the sham optrode into a cadaveric cat cochlea.
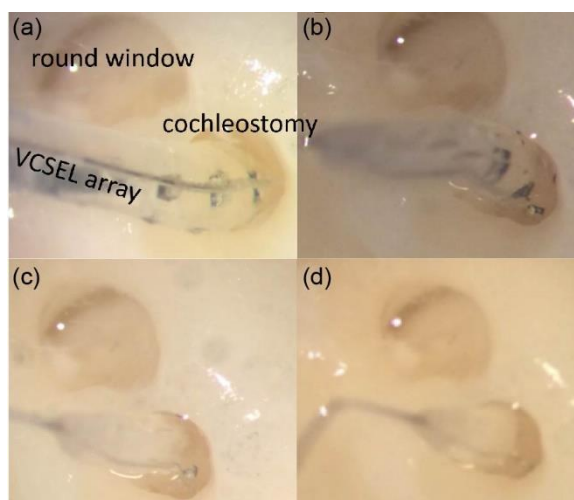
Figure 2-5. Panel (a) shows the landmarks of the magnified view of the cochlea, the round window, the cochleostomy, and the optrode. (b) through (d) show the progressive insertion of the optrode into the cochlea. In panel (d) the entire optrode is inserted into scala tympani. The length of insertion is about a 6 mm. Considering the spacing of the optical sources and insertion depth, the maximum number of VCSELs that can be inserted into the cat cochlear at this time is five.

2.4.2 Insertion force measurements

Insertion force was measured with different custom-made optrodes in cat and human cochleae. For the short and thin arrays, the insertion forces were relatively small (<5 mN, Table 2-1). Insertion force was also tested with a plexiglass model of the human scala tympani, which allowed a direct view of the insertion depth (Figure 2-6). Four arrays were tested in this model: two custom-made optrodes and two electrical arrays from contemporary cochlear implant systems (Table 2-1, Figure 2-6). The insertion force and depth were comparable between the optrode with

15 contacts and the 16-channel electrode. Of note, the insertion depth was shorter than the entire length of the optrodes and electrodes. All optrodes and electrodes could be inserted with full length by hand (Figure 2-2(g)-(i)).

Table 2- 1 Insertion force measurements for optical and electrical arrays.

Type	Cochlea	Contacts No.	length (µm)	Width (µm)	Height (µm)	Insertion force (mN)	Insertion distance (µm)
µLED (blue)	Cat, Human	4	5000	920	710	<5	2686
µLED (blue)	Cat, Human	4	5000	790	620	<5	2965
µLED (blue) VCSEL (red)	Cat, Human	5	6000	790	570	<5	3286
VCSEL (infrared)	Cat, Human	3	4000	850	440	<5	2926
electric	Cat, Human	3	4000	780	540	<0.5	3552

μLED (blue)	Human model	10	15000	1100-1000	900-600	<150	8275
μLED (blue)	Human model	15	24000	1100-600	900-500	<130	13230
electric (commercial)	Human model	12	26400	1100-600	1100-600	<75	16530
electric (commercial)	Human model	16	24500	1100-600	600-400	<130	13600

Note: The table describes the different arrays used for the insertion force measurements in a cat cochlea, human cochleae, and a human scala tympani model. The number of optical or electrical sources, the width, the height, and insertion force and distance are also listed. The four arrays equal or longer than 15 mm decrease their width and height from base to apex. Changes are given by the ranges in the table.

Figure 2- 6. Insertion force measured in a model of the human scala tympani.
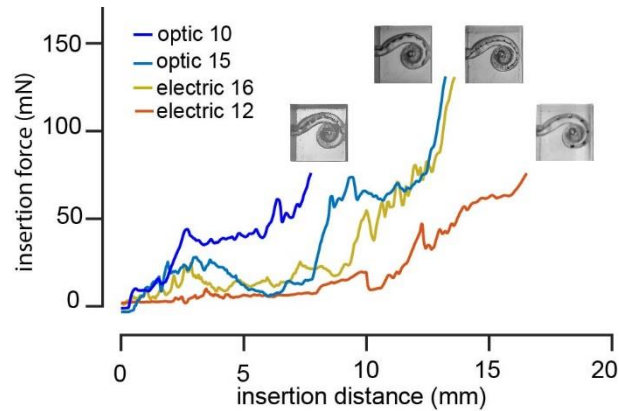
Figure 2-6. The figure shows the changes in force during insertion in a model of the human scala tympani at different depths of insertion for four arrays. The optrode has 10 (blue line) or 15 (turquoise line) µLEDs. The electrical alone arrays have 16 or 12 contacts. The placement of each array in the model is shown in the four corresponding inserts.

The placement of the optrodes in the cat cochleae was also examined by X-ray micro tomography with synchrotron radiation. Figure 2-7(a) shows a typical projection obtained during the scans. The reconstruction of the optrode is shown in Figure 2-7(b) and (d). Corresponding sketches are shown in Figure 2-7 (c) and (e).

Figure 2- 7. Placement of the optrodes in the cat cochleae.
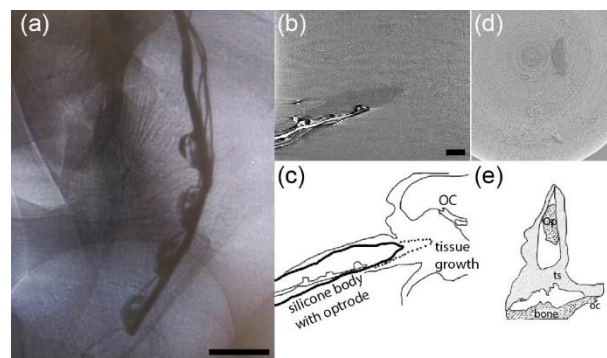
Figure 2-7. Panel (a) shows an X-ray projection of an inserted optical array *in situ* in a cat cochlea. The thick wire is the backbone and acts as a heat sink. The thin wires connect to the anodes of the optical sources. The scale bar represents 500 µm. Panels (b) and (c) show the same array after the reconstruction and its sketch. The optical sources irradiate Rosenthal's canal. A thin layer of tissue can be seen around the electrode, which has been slightly retracted (dash line in the sketch). The organ of Corti (OC) is also marked. The scale bar represents 500 µm and is used for the following panels. Panel (d) and (e) shows the cross section of the same array after the reconstruction and its sketch. The optrode (Op), tissue, bone, and OC are marked in the sketch.

2.4.3 In vivo functional testing in the guinea pig animal model

SELDs were assembled into arrays which could be inserted into the guinea pig cochlea for *in vivo* functional testing. At an input current of 600 mA, the maximum radiant energy emitted from the SELDs was 15.3 ± 4.9 µJ per pulse (n=10), ranging from 8 to 20 µJ per pulse. The energy of each SELD will be screened in future assemblies of the optrodes to eliminate the variation. The energy was measured in air with a Coherent J-50-LP-1A energy sensor. The SELDs arrays were inserted through the cochleostomy to stimulate the base of the cochlea. Optical pulses were 100 µs in duration and were delivered at 100 Hz. During the experiments, the test current levels were increased in 6 equal steps from 0 to 600 mA (Figure 2-8(a)). The amplitude of the CAPs dropped with decreasing current amplitude and reached a minimum at 300 mA (Figure 2-8(a)). CAPs also disappeared when the SELD arrays were rotated in or partially extracted from the cochleostomy so that the emitting side was no longer facing the spiral ganglion neurons.

Figure 2- 8. The traces show CAPs evoked with an SELD and an optical fiber in the same animal (guinea pig).

(a)

20 µJ

5 µV

Time (ms)

(b)

5 µV

27.8 µJ

16.9 µJ

15.2 µJ

14.2 µJ

Time (ms)

(c)

5 µV

laser fiber
SELD

Radiant energy (µJ/pulse)
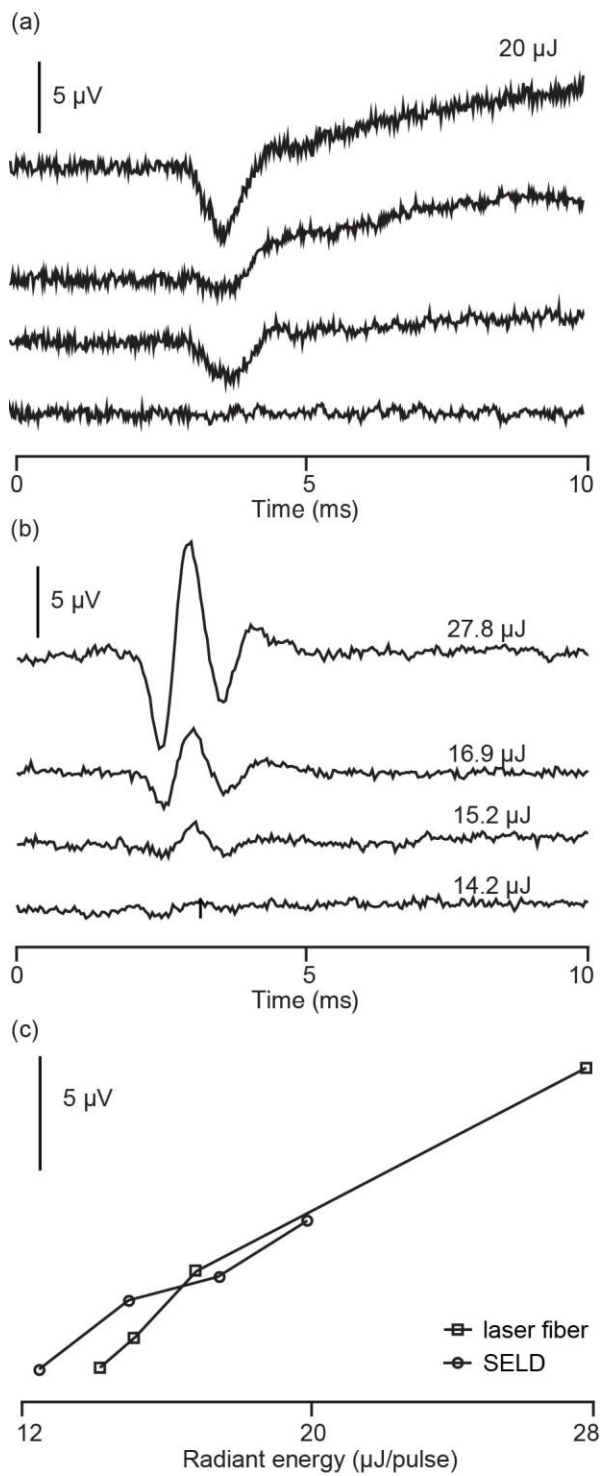
Figure 2-8. (a) shows CAP responses evoked by a SELD operated at different current levels. The energy output measured prior to the in vivo test was 20 µJ/pulse with 600 mA current input. The four traces represent the current input of 600, 500, 400, and 300 mA. The traces are the averaged responses to 20 stimulus presentations. (b) shows CAPs evoked by delivering the radiant energy with an optical fiber at different energy levels. The traces are the averaged responses to 100 stimulus presentations. (c) shows the CAP amplitudes at different radiant energies for both the SELD and the optical fiber.

CAP amplitudes in response to INS which were obtained with the SELDs were compared to CAP amplitudes measured in the same animal by delivering the radiation with an optical fiber (Figure 2-8(b), (c)) coupled to a table-top laser (Lockheed Martin Aculight Corp., Bothell, WA).

2.4.4 In vivo test in cats

For the novel cochlear stimulation devices, it is important that they are both biocompatible and can evoke neural responses. In cat cochleae, the power of the VCSEL array was too low to evoke measurable auditory brainstem responses. The results of the chronic experiments demonstrate that the optrodes are biocompatible and that chronic implantation does not damage cochlear function over time. Functional tests with more powerful SELDs were done in the guinea pigs as shown previously.

After the *in vitro* long-term testing, the optrodes were implanted into the left cochleae of a cat. Figure 2-4, 5 show the optrode insertion and the transcutaneous connector to the current source. The current sources were small computer controllable laser diode drivers developed by Lockheed

Martin Aculight (LMA), which fit into the backpack of a cat. Alternatively, a commercially available High-Power Precision Source, LDX-32400 (ILX Lightware, Bozeman, Mo), was used. Optically and acoustically evoked ABRs were recorded before implantation (baseline) and two weeks after the surgery (Figure 2-9). Thereafter, acoustically evoked ABRs were measured every two weeks up to 26 weeks after the surgery. Note, the cat study aimed to determine the biocompatibility of the optrodes and to determine whether implantation and materials will lead to a deterioration of cochlear function over time. Cat number 13IKB3 was implanted with red VCSELs. Cats 13IMR3 and 13CKC6 were implanted with the low power infrared VCSELs. 13IKB3 was excluded from measuring the optical ABRs (oABRs). No oABRs were evoked by the low power VCSELs. In pulsed operation mode, the highest output power of low power VCSELs is about 70 mW (7 µJ/pulse), which is about the radiant energy required to reach stimulation threshold as determined in previous experiments [46, 178]. Sound levels to evoke an ABR with acoustic clicks were elevated immediately after surgery by about 50 dB, 35 dB, and 25 dB in cat 13IKB3, 13IMR3, and 13CKC6 respectively. Thresholds were determined two weeks after the surgery. No further changes in threshold were observed after the placement of the optrode (Figure 2-9(a)). Threshold elevations occurred mostly at frequencies above 22 kHz, where the optrode was located (Figure 2-9(b)). No acoustic responses could be recorded for stimulation frequencies above 22 kHz in all three cats, and sound levels to reach the threshold for an ABR were elevated at frequencies between 8 kHz to 22 kHz by 25 dB in cat 13IMR3.

Figure 2- 9. This figure shows the ABR thresholds to acoustic stimuli pre- and post-implantation.
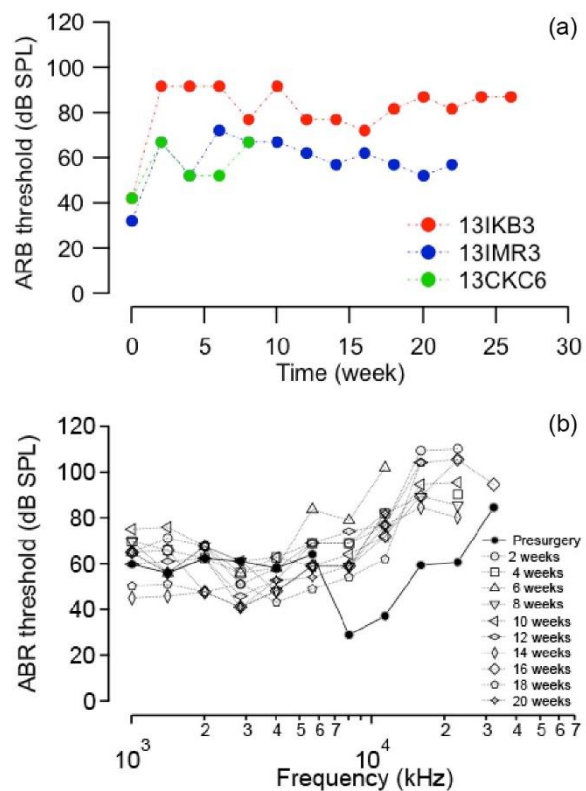
Figure 2-9. (a) shows Click evoked ABR thresholds at different time points in three cats. Click thresholds were elevated after optrode implantation, but then remained consistent in the months after; (b) shows pure tone evoked ABR thresholds at different times after implantation for animal 13IMR3. The optrode implantation caused high frequency hearing loss (at 32 kHz) and elevated thresholds among 8 kHz to 22.6 kHz, but little change was noted below 8 kHz.

2.5 Discussion

2.5.1 Requirements for the LDS

Physical design of the light delivery system

In the ideal scenario, the cochlear implant electrode design would be tailored to each patient. This would allow the optimal placement and orientation of the optical sources towards the auditory neurons. Since this technology has not matured yet, the electrode array was fabricated in a circular form to best fit the different area configurations of scala tympani. Ample data are available in the literature describing the dimensions of scala tympani of the human cochlea [77, 179-181]. The dimensions of the LDS should taper from 1 mm to 0.47 mm. The length of the electrode/optrode should be shorter than 27 mm.

Energy requirements

In a recent paper, we reported the radiant energy required for INS in guinea pigs and cats [77]. At the tip of the optical fiber, it was on average 14.1 ± 8.1 µJ/pulse for single units in the central nucleus of the inferior colliculus (ICC) and 17.2 ± 13.9 µJ/pulse for compound action potentials (CAPs). The variation of the radiant energy at the tip of the optical fiber to reach stimulation threshold was large with a range of 4.8 to 47 µJ/pulse. After correcting for the distance between the tip of the optical fiber and the modiolus, the radiant energy on target was between 1.4 and 16.4 µJ/pulse, on average 4.1 ± 1.9 µJ/pulse for ICC single units and 7.2 ± 4.7 µJ/pulse for CAP responses. The radiant energy of the SELDs emitting light at 1850 nm was typically in the range between 8 and 20 µJ/pulse. This reaches the threshold for stimulation in most of the cases. It is also below the threshold for which cochlear damage was detected [182, 183]. It is important to explore and implement methods to reduce the amount of energy required for stimulation. One possibility is combined optical and electrical stimulation. Duke et al. have demonstrated that sub-

threshold electrical stimulation can lower the threshold for INS by a factor of about two [184, 185]. More recently, we have shown in deaf white cats that combined optical and electrical stimulation reduces the threshold for INS in the cochlea [185].

2.5.2 Enabling technology

Optical fibers and fiber bundles

A detailed study on the design of a light delivery system using optical fibers was completed previously [186]. To accommodate the optical fiber bundle in scala tympani of a human inner ear, the maximum diameter of the optical fiber bundle must be less than 0.7 mm. In their experiments, the authors inserted single silicone coated fibers or small silicone coated fiber bundles with core/cladding diameters of 20/25, 50/55, 50/125, and 105/125 µm. The results showed that thicker fibers (50/125 or 105/125) broke after being inserted about 10 mm. At this insertion depth, the tip of the electrode reaches a site along the cochlea where a steep turn occurs. Thin optical fibers and their corresponding optical fiber bundles up to 8 fibers could be inserted up to 20 mm into scala tympani of a cadaveric human cochlea. The results also demonstrate that optical fibers pose a challenge if more than 8 channels are required. Moreover, it is not clear whether the maximal energy delivered through those fibers is sufficient for stimulation. Based on their experience and working with glass optical fiber, we have not attempted to build a multi-channel optical implant to be deeply inserted into a cochlea.

Bundles of waveguides

Waveguides are typically made out of dielectric materials. Their elastic modulus is about 3.2 GPa versus ~70 GPa for fused silica. Depending on the geometry of the fiber's cross section,

waveguides are about 20-30 times more compliant than optical fibers made of fused silica. Waveguides core structure has a high index of refraction and is surrounded by a material with lower permittivity, the cladding. The structure guides optical waves by total internal reflection. Consequently, the extinction coefficient for the core material must be low for the radiation wavelength. While waveguides are readily available for the visible range, they are not for the near infrared or infrared. According to the literature, materials exist which are suitable for insertion into the cochlea and have good transmission at the wavelength of interest, λ=1840-2100 nm [187-189]. Examples are polyimides, such as Kapton®. It is commercially available through DuPont. The index of refraction of the material at λ=1860 nm is about 1.6. At those wavelengths, Kapton® waveguides (size 10 µm) have propagation losses around 1 dB/cm. Fluorinated polyimides are even better at transmitting the radiation. They have propagation losses of about 0.6 dB/cm [190]. The cladding of the waveguides could be Teflon or Silastic®. The latter is used to encapsulate the waveguides. Silastic® and Teflon have an index of refraction of about 1.3 at λ=1860 nm. This gives a critical angle of approximately 60 degrees.

Optogenetics

For the mouse, it has been demonstrated that the blue light radiation at 2.2 µJ/mm^2 evokes an auditory response. This is about 7-70 times less than the energy required for INS in the gerbil or cat. Existing µLEDS are powerful enough for stimulation. The technology has advanced to produce miniature sources [191-193] that can be inserted in cochleae of small animals. However, challenges are expected for larger distances from the light sources and bone lies in the beam path. As indicated before, for radiation wavelengths up to about 1064 nm, the extinction coefficient for

the radiation is governed by the scattering of the photons. For wavelengths above 1064 nm, the absorption of the photons by the fluids in the tissue becomes the dominant factor. In other words, bone in the beam path will drastically reduce the radiant energy and will widen the beam path and consequently will affect the selectivity of optical stimulation.

2.6 Conclusion

With our work, we have demonstrated that light delivery systems can be fabricated for optical stimulation with INS. With the current design, optrodes can be inserted into a human scala tympani model up to approximately 360 degrees. Optrodes made with SELDs were able to evoke auditory responses in guinea pigs. Chronic implantation of the optrodes did not elevate acoustically evoked auditory brainstem responses over 26 weeks in cats. Future studies will focus on developing optrodes with SELDs and testing functionality in longitudinal studies.

2.7 Acknowledgments

2.8 Disclosures

The authors have declared that no conflict of interest exists. US Patent No. 7833257 entitled "Optical stimulation of the auditory nerve, a novel concept for cochlear implants"; inventors: Walsh Jr., J., Izzo, A., Jansen, D., Richter, C.-P. (2010) has been issued to Northwestern University. No licensing arrangements for this patent or any other interests of a financial nature currently exist.

Chapter 3:

Distorting temporal fine structure by phase shifting and its effects on speech intelligibility and neural phase locking

Yingyue Xu[a*], Maxin Chen[b*], Petrina LaFaire[a], Xiaodong Tan[a], Claus-Peter Richter[a-c]

[a] Northwestern University, Department of Otolaryngology, 320 E. Superior Street, Searle 12-561, Chicago, IL 60611, USA.

[b] Northwestern University, Department of Biomedical Engineering, 2145 Sheridan Road, Tech E310, Evanston, IL 60208, USA.

[c] Northwestern University, The Hugh Knowles Center, Department of Communication Sciences and Disorders, 2240 Campus Drive, Evanston, IL 60208, USA.

* authors equally contributed to the study

This chapter was published in:

"Distorting temporal fine structure by phase shifting and its effects on speech intelligibility and neural phase locking,"

Scientific Reports. 7, Article number: 13387 (2017)

3.1 Abstract

Envelope (E) and temporal fine structure (TFS) are important features of acoustic signals and their corresponding perceptual function has been investigated with various listening tasks. To further understand the underlying neural processing of TFS, experiments in humans and animals were conducted to demonstrate the effects of modifying the TFS in natural speech sentences on both speech recognition and neural coding. The TFS of natural speech sentences was modified by distorting the phase and maintaining the magnitude. Speech intelligibility was then tested for normal-hearing listeners using the intact and reconstructed sentences presented in quiet and against background noise. Sentences with modified TFS were then used to evoke neural activity in auditory neurons of the inferior colliculus in guinea pigs. Our study demonstrated that speech intelligibility in humans relied on the periodic cues of speech TFS in both quiet and noisy listening conditions. Furthermore, recordings of neural activity from the guinea pig inferior colliculus have shown that individual auditory neurons exhibit phase locking patterns to the periodic cues of speech TFS that disappear when reconstructed sounds do not show periodic patterns anymore. Thus, the periodic cues of TFS are essential for speech intelligibility and are encoded in auditory neurons by phase locking.

3.2 Introduction

Speech is a robust signal that remains intelligible despite various means of perturbation[193]. Key components necessary for speech intelligibility have been examined. For example, speech containing only the envelope information (E, slow varying components <50 Hz) or the temporal fine structure information (TFS, rapid varying components >50 Hz) of a few

frequency bands was found to be intelligible for normal hearing listeners[110, 194, 195]. The role of TFS has also been extensively examined through other hearing tasks. These tasks have shown that TFS is essential for speech intelligibility in noise, music perception, sound localization, and frequency discrimination. To demonstrate this, the low frequency spectrum (50-500 Hz) of TFS has been isolated and defined as conveying periodic cues by some researchers, also known as the temporal periodicity[91-96]. These periodic cues were specifically examined since they are abundant in human speech, music and even non-animated sounds due to physical constrains[92]. Findings from these experiments suggested that periodic cues are important for the perception of rhythm and syllabicity[91, 196], pitch, voicing, intonation, sound localization, and identifying a common sound source that activates different frequency bands under noise[92, 93, 113, 115, 197, 198].

The significance of TFS for speech intelligibility has been demonstrated through human behavior studies[94]. Studies have also been carried out to examine the underlying neural mechanisms related to TFS, specifically the periodic cues. Recent studies, which used Magneto- and electroencephalography showed clear representations of TFS in the recorded cortical activity[95, 96]. Other studies attempted to assess the internal representation of the TFS using physiologically-plausible models of the auditory periphery[199-202]. More direct and fundamental neural correlates to the TFS cues have been explored in animals by measuring the discharge patterns of auditory neurons, in particular the timing of neural discharges or the phase locking of the neurons[86]. Phase locking is a phenomenon that is preserved along the auditory pathway from the cochlea to the cortex[97-105]. It is referred to as the tendency for auditory neurons to fire within a well-defined time window relative to a period of the stimulating frequency.

This is particularly evident in the neural response to low-frequency pure tones[86]. Phase locking pattern is also seen in auditory nerve fibers in response to periodic sounds, such as modulated frequencies and steady-state vowels. Importantly, the TFS of natural speech sentences contains multiple representative frequencies, including the fundamental frequency (F_0) and its harmonics. Auditory neurons may discharge at the rate of these frequencies[97, 106-108], i.e., phase lock to these frequencies. Thus, the phase locking pattern could serve as a tool in examining the neural coding of the TFS in speech.

While phase locking could be used to investigate the neural components of TFS, reconstructed speech sounds could be used to investigate its perceptual functions. Previous studies have frequently manipulated speech sounds to assess the perceptual role of different acoustic cues. Modified speech could be achieved by keeping the acoustic cues of interest, while disrupting or replacing other acoustic components, and vice versa. For example, studies examining E have reconstructed acoustic signals by replacing spectral cues with modulated sinusoids or pulse trains[109] or spectrally matched noise[110]. The E is usually extracted by a Hilbert transformation or by the low pass filtering of the rectified bandpass filtered speech signal. Through dividing the E from the original signal or peak clipping, TFS cues could be be isolated. An alternative method to distort speech is based on the short-time Fourier transform (STFT)[111] or Wavelet transformation[112], which generates the magnitude and phase spectrogram. Distorting either the magnitude or the phase spectrogram would affect the intelligibility of speech. Therefore, a systematic and targeted distortion of TFS cues could be achieved by systematically distorting the phase spectrogram while maintaining the magnitude spectrogram. To further examine the neural mechanism underlying TFS processing for speech perception, we adopted the STFT approach to

systematically distort the TFS cues in speech. The magnitude and phase spectrogram of natural speech was calculated with fine time resolution. The phase was shifted to various extents for sentence reconstruction. The reconstructed sentences featured systematic distortion in TFS and were used to (1) test speech perception in normal hearing listeners in both quiet and noisy listening conditions and (2) test neural activities in the auditory neurons in animals.

Our study focused on the neural mechanism of TFS processing by combining neural recordings from individual auditory neurons in guinea pigs with psychophysical measures in human subjects. The TFS cues of sentences from the Hearing in Noise Test (HINT) and the QuickSIN™ test were modified to various extents. The modified speech signals were used for speech perception tests in normal hearing listeners in both quiet and noisy conditions. These same sentences were used to evoke neural activity in well-identified auditory neurons of the inferior colliculus (ICC) in guinea pigs. Through this direct comparison between behavioral and neural data, our results bridged the perceptual role and neural coding of TFS cues in speech, allowing us to uncover the essential neural mechanism for TFS perception.

3.3 Results

3.3.1 TFS cues were distorted by phase shifting

Speech sentences with altered TFS cues were designed with little distortion of the speech E (see Methods for details). For the animal experiments, the original speech sentences (no noise presented) were selected from the HINT test and referred to as sentence type 1 (S1). A 64-channel spectrogram of each original speech sentence (S1) was calculated and the six frequency bands with

the maximum energy were then selected for sentence reconstruction. Sentences reconstructed with the 6 frequency bands selected from the original S1 were referred as sentence type 2 (S2). To alter the TFS, the phase information within the six selected channels was distorted to various extents. Phase information was randomly shifted within half a period to reconstruct sentence type 3 (S3). Phase information was randomly shifted within a whole period to reconstruct sentence type 4 (S4). Phase information was also replaced by a random number to reconstruct sentence type 5 (S5). Phase information was set to zero to reconstruct sentence type 6 (S6). Finally, the magnitude of each sentence type was normalized to the magnitude of the original spectrogram.

This phase distortion severely degraded the TFS, but not the E of the speech. To illustrate this point, a sentence from the HINT test was used as an example. This sentence was "the silly boy's hiding" and it was 2 seconds long. Each sentence type, S1 - S6, can be compared as a function of frequency and time via its corresponding spectrogram (Figure 3-1(a)). Visual inspection of the spectrograms showed relatively similar magnitudes, envelopes, and spectral contents. Furthermore, the spectrograms of S2   S6 were strongly correlated with S1, with the Pearson product-moment correlation coefficient (CC) index above 0.7 (Figure 3-1(c)). Namely, the E was well preserved from S1 to S2   S6. This was supported by the neural histograms (see Supplementary Figure 3-S1 and S2 online), which showed comparable responses to all six speech signals, demonstrating that the E of S1 - S6 was processed similarly in auditory neurons.

To further examine the TFS, sentences were divided into smaller time sections, for example, the second "i" from the sentence "the silly boy's hiding". A Fourier transformation was applied to this voiced syllable in each sentence type, S1 - S6. The spectrum for "i" in sentence S1 showed

three large representative peaks in the frequency range from 50 to 500 Hz, i.e. the F_0, the 1st and 2nd harmonic of F_0. These TFS cues seen in the S1 spectrum were distorted to different degrees in S2 S6, particularly affecting F_0, the 1st and 2nd harmonic of F_0, i.e. the periodic part of TFS of this syllable (Figure 3-1b). This dominating temporal periodic pattern was generally preserved in S2, i.e. the representative frequency peaks were observed in the S2 spectra with similar magnitudes. In other words, sentences reconstrued with fewer frequency bands did not degrade the periodic cues of TFS. This periodic pattern was also seen in S3, though the magnitudes of the peaks were slightly decreased. This indicated that phase shifting within half a period didn't eliminate the periodic cues of TFS. The TFS patterns were further diminished in S4, S5, and S6. Apart from these representative frequency peaks, the spectrograms of S1 to S5 were well matched in a contiguous spectral regions [110]. The CC index was calculated between the original and altered speech signals for each of the selected voiced syllables from the speech sentences. For example, there were nine selected voiced syllables "th, e, i, lly, b, oy, i, d, ing" in the sentence "the silly boy's hiding". The averaged CC index across the nine syllables was calculated, as well as the standard deviation (Figure 3-1(c)). S2 and S3 were strongly correlated with S1, with an average CC index above 0.7. However, S4, S5, and S6 showed weak correlation to the original speech, with an average CC index around 0.2.

Figure 3- 1. An example of speech stimuli with different levels of phase distortion.
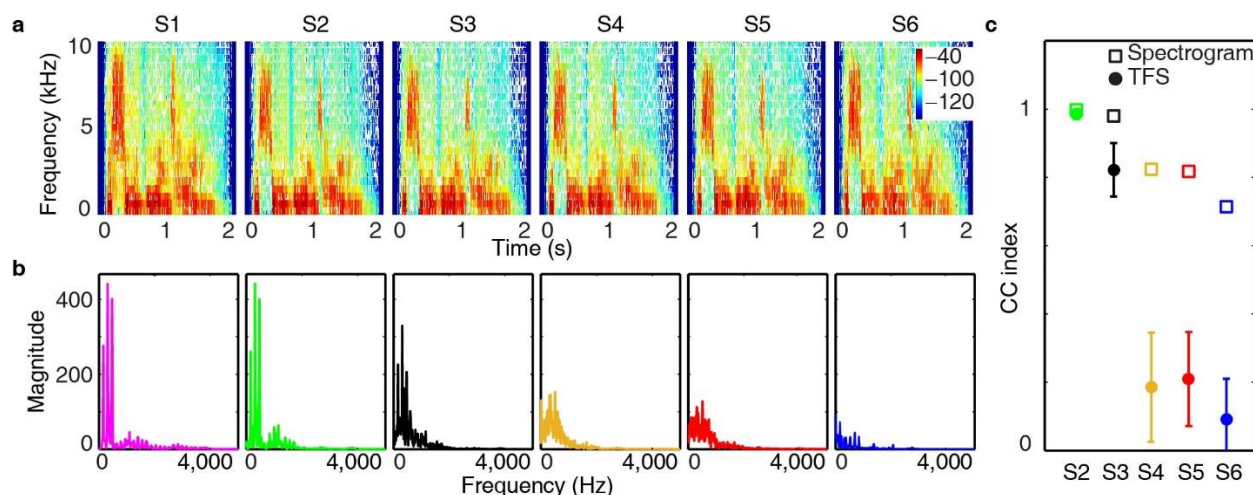
Figure 3-1. The narrow band spectrogram of S1 to S6 shows the energy of the sentence as a function of time and frequency (a). The CC index was calculated between the spectrogram of S1 and S2   S6 (c). The second "i" from the sentence "the silly boy's hiding" was selected as an example. The spectra of "i" from 0 to 5000 Hz were plotted for each type of sentence (b). Similar color coding was used for all following figures. The representative peaks in S1, S2, and S3 share the same frequencies, although the magnitudes were slightly different. For each voiced syllable, the CC index was calculated between the spectra of S1 and S2 – S6. The CC index was averaged across all voiced syllables (n = 9) in the sentence. The averaged CC index and the standard deviation were plotted for each sentence as solid circles and error bars (c). The CC index was also calculated between the spectrogram of S1 and S2 – S6, as shown in open squares (c).

3.3.2 Speech perception was affected by TFS in quiet and noise

Speech intelligibility of the altered sentences was examined in normal hearing subjects using sentence lists from the QuickSIN™ test, both in quiet listening environments and with background noise. The original sentence, S1, was not tested since subjects were normal hearing listeners. The results of the speech test for S2 – S6 in quiet were plotted as a box plot, as shown in Figure 3-2(a). Speech intelligibility was near perfect for S2 (with reduced number of frequency bands). Speech intelligibility was over 80% on average for S3 (phase shifted within half a period). However, speech sentences became hardly intelligible for S4 (phase shifted within a full period), S5 (phase replaced by a random number), and S6 (phase replaced by 0). The average scores for correct answers and the corresponding standard deviations were 87% 4% for S2, 81% 8% for S3 23% 7% for S4, 17% 10% for S5 and 14% 8% for S6. Speech performance decreased systematically with TFS distortion. Differences in outcomes were evaluated between neighboring sentences using Wilcoxon signed rank test with 95% confidence interval. The p-value at 9 degree of freedom is taken as the standard for statistical analysis. A p-value smaller than 0.05 was considered statistically different between conditions. The p-values were 0.061 between S2 and S3, 0.002 between S3 and S4, 0.049 between S4 and S5, 0.193 between S5 and S6. The significant difference between speech score for S3 and S4 suggested that speech perception in the present experiments was affected by the distortion in the TFS.

Speech recognition was also tested with sentences from the QuickSIN™ test for noisy listening conditions with five different signal-to-noise ratios (SNRs), ranging from 5 dB to 25 dB in 5 dB steps (details seen in Methods). The averaged speech recognition scores and standard deviations were calculated (Figure 3-2(b)) for five speech sentences (S2 – S6). Speech performance decreased with the reduction of SNR for all tested sentence types. The speech

recognition scores also clustered into two groups: S2 and S3, as one group with overall high scores for all SNRs, and S4 to S6, as another group with poor speech scores. The separation of the two groups clearly showed that speech recognition in noisy listening environments was superior when the TFS was preserved: S2 and S3 were better recognized than S4, S5, and S6. With an SNR of 5 and 10 dB, the speech signal was almost unintelligible for S4, S5, and S6.

Figure 3- 2. Speech recognition was tested in normal hearing subjects with five types of reconstructed sentences, S2 to S6.
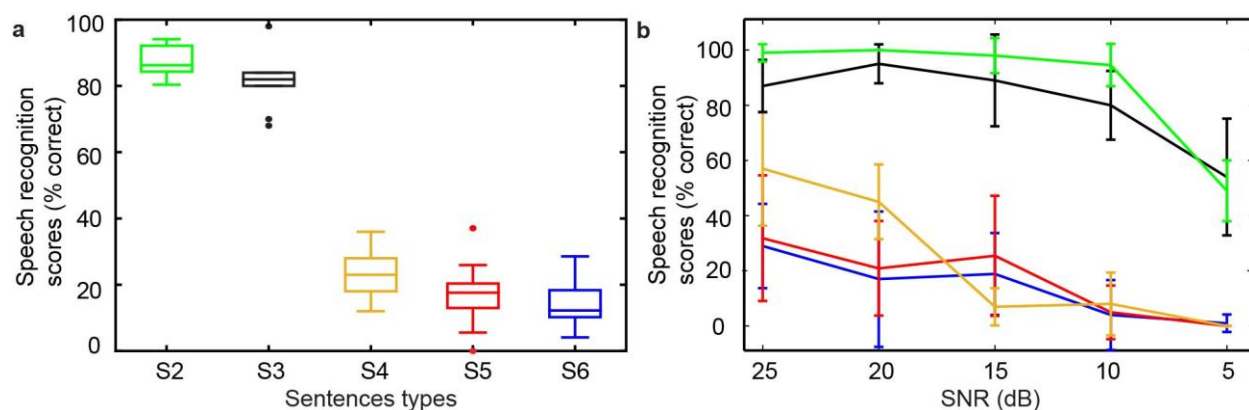
Figure 3-2. The scores for all test subjects for speech in quiet were shown in the box-and-whisker plot (a). The maximum, median, minimum, first and third quartiles were also shown in the plot. Dots outside whisker boxes were outliers not included in according boxes, but still considered in comparison. Speech recognition scores were also obtained for speech in noise for five different SNRs: average scores and standard deviations are plotted for all subjects (b). Five types of speech files were presented in a random order for each condition to alleviate possible training effects or tiredness of the subjects from the results.

3.3.3 Neural phase locking to TFS

The coding of TFS cues in the speech signal was also examined at the level of individual auditory neurons. The neural activity in response to all six types of sentences was recorded from well-defined single units in the ICC of guinea pigs. Neural coding of the period cues of TFS was examined by neural phase locking[91]. The phase locking patterns to the TFS cues could be derived from the spectral analysis of the post-stimulus time histogram (PSTH) (for details see Methods). A voiced syllable "o" from "the silly boy is hiding" was selected: time section from 0.714 to 0.914 s post the onset of each sentence and PSTH. A Fourier transformation was applied to the neural PSTHs to extract the phase locking pattern. The spectra of the PSTHs evoked by S1 and S2 (Figure 3-3(a), (b)) contain three large peaks at 105, 210, 315 Hz respectively. These large peaks indicated that the neural discharge was phase locked to these frequencies. These frequencies corresponded to the frequencies seen in the FFT of the acoustic signal: the F_0, as well as the 1st and 2nd harmonic of F_0. Neural PSTHs showed identifiable peaks at 420 and 525 Hz with relatively smaller magnitude when compared with the first three peaks in the spectra of S1 and S2. The PSTH evoked by S3 also contained peaks at 105, 210, 315 Hz, but not at 420 and 525 Hz (Figure 3-3(c)). Despite selecting only six frequency bands (S2) and shifting the phase within half a period (S3), the speech signals still showed the dominant frequency peaks seen in the original sentence (S1). Thus, the frequencies within the TFS cues were well preserved from S1 to S2 and S3, yielding phase locking in the auditory neurons, as well as normal speech recognition in humans (Figure 3-2).

On the contrary, in response to S4, S5, and S6, the spectra of neural PSTH did not show any large representative frequency. Namely, no robust phase locking was seen (Figure 3-3 (d) to (f)), which was likely due to the absence of the periodic cues in the TFS resulting from altering the phase within a whole period (S4), or more (S5 and S6). The disruption of the periodic pattern in the TFS resulted in the absence of neural phase locking and speech recognition in humans (Figure 3-2). In other words, the periodic pattern in the TFS of speech was encoded by the auditory neurons through phase locking.

Figure 3- 3. Spectra of the acoustic stimuli and neural PSTH.
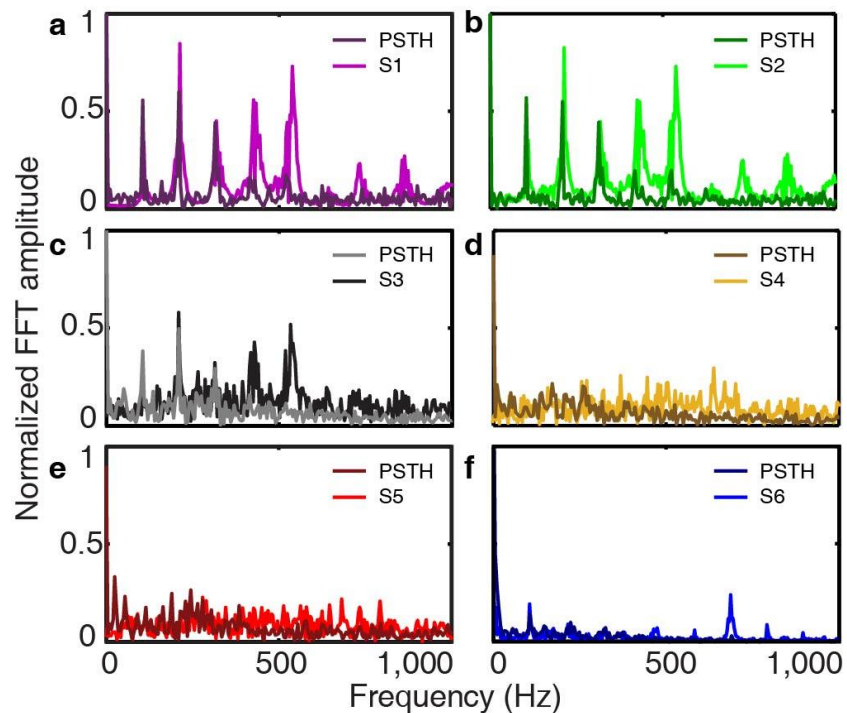
Figure 3-3. Spectra of the acoustic stimuli were plotted along with the spectra of their corresponding PSTHs calculated from the neural activity obtained during acoustic stimulation

from one representative unit in the guinea pig ICC. The best frequency (BF) of the unit was 600 Hz.
A voiced syllable "o" from "the silly boy is hiding" was selected (0.714 to 0.914 s post the onset
of each sentence). S1 to S6 were shown in a to f. The magnitude of spectra was normalized to the
maximal magnitude with 1000 Hz to allow a direct comparison. The first three dominating peaks
in S1 (a), S2 (b) and S3 (c) were at 105, 210, 315 Hz respectively.

Similar results have been observed in response to other voiced syllables, as well as in other
ICC neurons with different BFs. Phase locking was quantified by the synchronization index (SI),
i.e., the ratio between the magnitude of the frequency that neurons phase locked to and the total
number of discharges recorded[106, 118, 123]. For each sentence type, the SI was calculated at
the first three frequencies of the given syllable. The SI was averaged over 36 individual neurons
(Figure 3-4(a)). SI values at the F_0 were comparable for S1, S2, and S3, but dropped significantly
for S4, S5, and S6, indicating a loss of phase locking. Differences in outcomes were tested with
the Student's t-test and were statistically significant for S3 and S4 at the F_0 (p = 0.0005). The CC
between the spectra of PSTHs and speech signals was also calculated for all 36 single units (Figure
3-4(b)). A systematic decrease in average CC was shown from S1 to S5. S6 showed a relatively
large CC, which might be because the F_0 was still visible, although the magnitude was largely
decreased compared to the original sentence, S1 (Figure 3-3(f)). This small F_0 induced a small
phase locking response, resulting in a large CC, as well as a relatively large variation in the SI
values at this frequency (Figure 3-4(a)).

Figure 3- 4. The average and standard deviation of the SI of 36 single units is shown for F_0 and the 1st and 2nd harmonic of the selected voiced syllable "o" from "the silly boy is hiding".
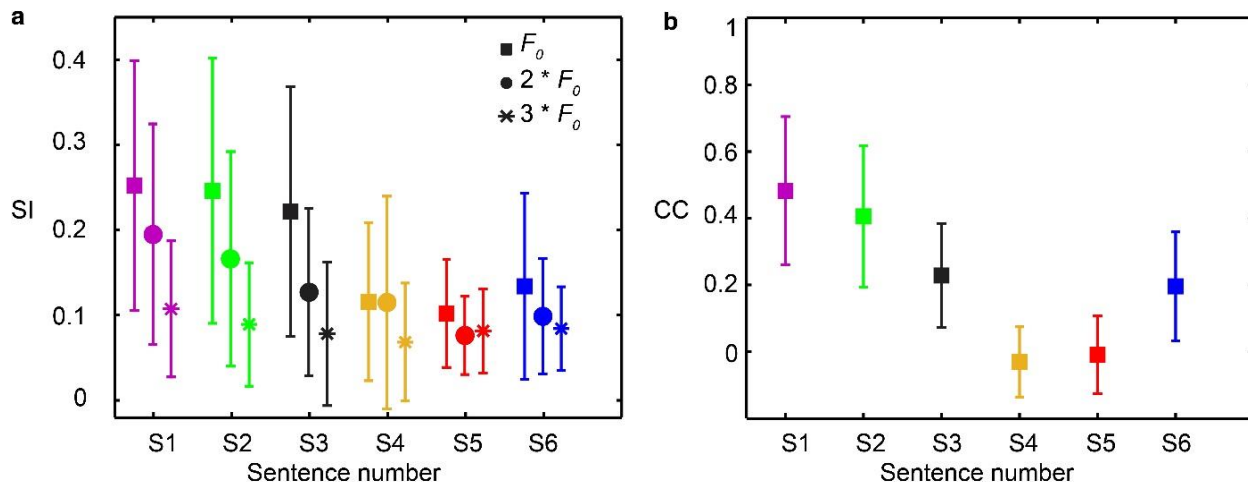
Figure 3-4. (a). The average and standard deviation of the CC between the spectra of neural PSTH from 36 single units and the corresponding acoustic spectra (b). The BF of these neurons ranged from 600 Hz to 3600 Hz.

3.4 Discussion

The results support the current view that phase locking is indeed an important factor to process TFS information. The results obtained from human testing have demonstrated that speech intelligibility in quiet could be deteriorated when the phase shifting range was a whole period, but could be maintained when the phase shifting was within half a period. Phase shifting within half a period also showed better understanding in noise. Corresponding experiments in guinea pigs have shown that individual auditory neurons exhibit phase locking patterns to the TFS in speech that disappear with the degradation of TFS in reconstructed sounds. Overall, speech remains

intelligible for a phase shift less than half a period (S3). Previous studies have shown that phase distortion compromised speech perception[111]. The results from this study demonstrated that some changes in TFS are tolerated before performance changes drastically. The maximum tolerable shift in phase was within half a period. In terms of acoustic properties, phase shifting within half a period maintained the periodic pattern of speech TFS. The preserved TFS also evoked phase locking in auditory neurons. This observation is consistent with the findings of neural phase locking. Auditory neurons tend to fire within a certain range relative to one sinusoidal cycle, not just to a single fixed time point of each cycle. The range is a quarter to half of a sinusoidal cycle[203]. Such localized firing within a time range could be beneficial for pitch detection of auditory neurons[204, 205]. It also indicates that a fixed rate of stimulation is not necessary when trying to include TFS in cochlear implant coding strategies. The results show, from both acoustic and neural processing perspectives, that a certain range of phase distortion, i.e. within half a period, does not affect speech intelligibility.

Sentences with TFS distortion were played to humans and guinea pigs for parallel comparison. Previous studies have demonstrated that the auditory tuning curves obtained from action potentials were comparable between humans and guinea pigs. Specifically, these studies showed that there is a general broadening trend from high to low frequencies[205-207]. The sharpness of tuning measured from action potentials was reported to be smaller than those obtained from single-fiber tuning curves in guinea pigs[208] and psychoacoustically in humans[209]. On the other hand, new otoacoustic measurements indicated that cochlea tuning in humans is sharper than other mammals, including guinea pigs[210]. To overcome the controversial differences of auditory tuning between humans and guinea pigs, the stimulation level was high to evoke broad

tuning for both species. The sound level used for human speech perception tests was 61 dB (re 20 µPa). The sound level used for neural activity recording was 80 dB (re 20 µPa). Thus, the auditory systems of humans and guinea pigs share similar broad width of tuning at these sound levels.

The spectrogram of S1 to S6 was comparable regarding magnitudes, envelopes, as well as relative spectral contents as demonstrated in Figure 3-1. However, there was a significant difference between the speech perception evoked by these sentences. The difference in speech intelligibility was consistent, with the distortion in periodic patterns of TFS. Previous studies have demonstrated that with only temporal cues of 4 frequency bands, speech was intelligible for normal hearing listener[110]. These temporal cues were extracted by a low-pass filter with cutoff frequencies at 16, 50, 160 or 500 Hz. The slope of the low-pass filter was 6 dB per octave. Thus, the temporal cues may contain energy of 50 to 500 Hz, i.e. the periodic cues of TFS including the F_0, as well as the 1st and 2nd harmonic. Some other studies have shown that speech with only the E (< 50 Hz) of 16 frequency bands was able to maintain intelligibility for normal hearing listeners in noisy condition[195, 211]. Although TFS may not be seen within each frequency bands, the large number of frequency bands may have compensated for the temporal loss through the place coding. However, our study showed that speech signals with periodic cues of only six channels out of 64 frequency bands could maintain intelligibility in noise. The small number of frequency bands was selected to mimic the n-of-m coding strategy for cochlear implants. The number of independent frequency bands for cochlear implant users is usually 4 to 8. Thus, cochlear implant users may benefit from a coding strategy containing periodic patterns of TFS in addition to the E.

Although not examined systematically, most test subjects (6 out of 10) could recognize words with strong consonantal syllables, such as "sh" and "s" from audio files, with large phase distortion under low or medium noise levels (>15 dB SNR), but hardly recognized vowels. This might be because neural phase locking is evoked mainly by the periodic pattern in vowels, but not for high frequency consonants. Thus, the periodicity distortion had smaller effects on consonantal syllables than on vowels. The zero phase condition (S6) gave a robot-like sound with less tonal information according to the subjects, which decreased the scores for speech recognition (Figure 3-2). This is consistent with previous findings that temporal periodicity is important for tonal and vocal information in audio signals[194].

During the test, some subjects (4 out of 10) stated that they could not tell the gender of the speaker for S4, S5, and S6 (phase shifted within one period or more). Typically, the female voice was identified as a male. This is consistent with previous studies showing that adding TFS cues could increase gender discrimination with limited frequency bands of signal processing[212]. Since gender discrimination is mainly determined by the F_0 perception[213, 214], having F_0 preserved in the speech might benefit gender discrimination. During the experiment, the test subjects reported that they grew accustomed with the speech sentences. This might result in an increase in speech intelligibility. The speech intelligibility tests were conducted firstly in the quiet environment, and then in noise (five levels of SNR). Thus, across the two types of speech tests (in quiet and in noise), the training effects might have caused the score to increase for S4 in noise at 25 dB SNR. On the other hand, within the two types of speech recognition test (in quiet and in noise), we have taken three approaches to minimize possible training effects: (1) 3 sets of practice speech files were presented to the subjects prior to the experiment to familiarize the subjects with

reconstructed speech. (2) For speech perception tests in quiet, the order of S2 to S6 was randomized to alleviate tiredness from sentences difficult to understand. (3) For speech perception tests in noise, both the order of S2 to S6 and the order of 5 SNR levels were randomized to alleviate learning effects. Across speech tests in quiet and noise, the learning effect was likely the reason behind the increased score for S4 in noise at 25 dB SNR. Within the speech tests in noise, the randomization of speech files minimized the learning effects so that it was not influential enough to alter the trend of the scores as SNR decreased. Additionally, with the randomization of speech files, there was a clear separation of speech scores: S2 and S3, as one group with overall high scores for all SNRs, and S4 to S6, as another group with poor speech scores.

Studies have suggested that listeners with sensorineural hearing loss may have an increase in the perceptual salience of envelope structure that could adversely affect speech perception in fluctuating background noise through loudness recruitment[214-216]. People with moderate flat hearing loss showed decreased ability to use TFS cues[195]. It has also been suggested that the temporal precision of E coding could be enhanced with hearing loss at equal stimulus sensation level in chinchillas[217]. In addition, pronounced distortions in the tonotopic coding of temporal cues in auditory nerve fibers have been found in chinchilla after noise-induced hearing loss[218]. Our study has shown that periodic cues of TFS are essential for speech perception in normal hearing listeners in both quiet and noisy condition. It would be interesting to examine the perceptual role and neural coding of periodic patterns in the TFS in subjects with different levels and types of hearing loss.

In summary, our study provides direct evidence that neural phase locking is the underlying mechanism for processing periodic patterns in the TFS, which is essential for speech recognition in quiet and noisy conditions. Shifting the phase spectrogram of STFT within half a period would partially maintain the periodic pattern of TFS in speech, yielding speech perception in normal human listeners and phase locking in guinea pig auditory neurons. Thus, preserving the partial periodic pattern of TFS could benefit speech intelligibility in noise, tonal perception and gender discrimination. Cochlear implant users may benefit from a coding strategy containing periodic patterns of TFS in addition to the E.

3.5 Methods

3.5.1 Signal processing

Speech sentences were processed through STFT and the phase information was manipulated to create acoustic stimuli with TFS distortion. The flow chart in Figure 3-5 shows the general signal processing procedure. An STFT was applied to the speech sentence to separate speech into contiguous frequency bands using MATLAB built-in function *spectrogram()*. The parameters for the STFT were selected that the output $S(f, t)$ was an n × m matrix of complex numbers with 64 frequency bands (n = 64), shown in Figure 3-5. The time window for STFT was around 1.5 ms, which mainly provide targeted TFS distortion below 689 Hz[111]. This frequency range was dominated by the periodic patterns of TFS.

Figure 3- 5. Flow-chart of signal reconstruction for TFS distortion. The input audio file was first analyzed by calculating the STFT.
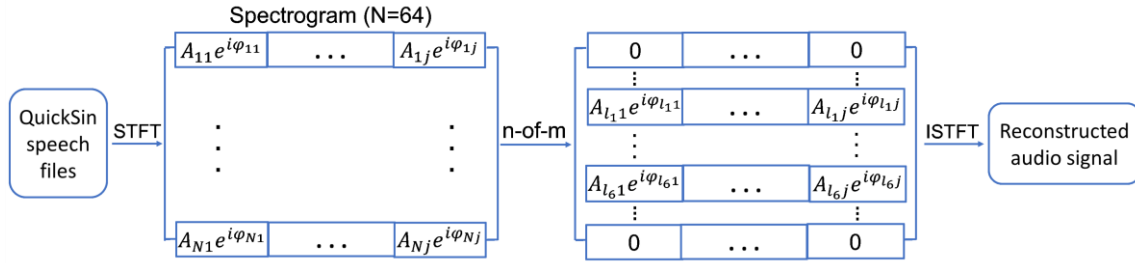
Figure 3-6. The resulting spectrogram contained 64 frequency bands. Applying the n-of-m method, the six frequency bands with highest energies were selected while eliminating other frequency bands to reconstruct new speech sentences. Magnitudes of selected channels were preserved, but phase values were modified under five conditions. The modified spectrogram was then used to reconstruct a new audio signal with phase distortion.

Each element $S(f, t)$ was converted into a combination of magnitude $|S(f, t)|$ and phase φ. To reconstruct speech sentences, the reverse STFT was applied. To mimic the coding strategy in cochlear implants, the number of frequency bands was reduced from 64 to 6. The frequency bands were selected using the n-of-m method: the six highest magnitudes were selected and reselected every 4 ms. All other frequency bands were eliminated. The 6 bands were then used to reconstruct speech sentences. To distort the periodic pattern, the phase values of the selected frequency bands were changed according to five conditions: 1) original phase values π_0, 2) random phase shift in a range of $[\pi_0 - \frac{\pi}{2}, \pi_0 + \frac{\pi}{2}]$, 3) random phase shifts in a range of $[\pi_0 - \pi, \pi_0 + \pi]$, 4) random numbers within $[-\pi, \pi]$, and 5) fixed numbers, zeros. These five conditions were used to reconstructed S2 to S6. With the original magnitude in the selected channels and the distorted phase values, equation 1

$$S(f,t) = |S(f,t)|e^{i\varphi} \qquad \text{(eq. 1)}$$

was used to reconstruct complex numbers in the new spectrogram S'. By calculating the inverse-STFT from the new spectrogram, S2 to S6 was reconstructed. The noise masker used for the speech in noise tests were 4-talker babble and was part of the track. Therefore, both the target speech and noise maskers were processed through phase distortion. For the sentences from the HINT test one channel contained the noise signal, the other channel the speech signal. For the speech in noise experiments no sentences form the HINT test were used.

3.5.2 Human test subjects and approach

Ethics statement:

The study was approved by the Northwestern Institutional Review Board (IRB), filed with the version number STU00201779. Informed consent was obtained from all participants at the time of their enrollment in the study.

Subjects:

For the experiments, 10 self-reported normal hearing native English-speaking subjects (3 females and 7 males, average age of 28 ± 8.8) were recruited. Normal hearing was confirmed by conducting a hearing test (see also supplementary information).

Sentence test:

The speech test was conducted in a quiet room with a noise level typically less than 40 dB (re 20 µPa). Speech audio files were played to subjects via calibrated headphones at 61 dB (re 20 µPa). A practice test with three sentences, each under a different phase distortion condition that subjects may come across in the test, were played first to the subjects so that they could familiarize themselves with the process. Next, ten new sets of sentences (two sets for each sentence types from S2 to S6) were played in a random order. Subjects were asked to repeat what they heard as accurately as possible after every sentence. Also, during the test, subjects were encouraged to describe their experience of the audio files, including pitch, timber and gender recognition.

The QuickSIN™ user manual provides a score sheet for each of the sentences, which was used for evaluation. Percentages of correct meaningful words repeated by subjects were calculated as the score for each sentence. To evaluate performance difference among different phase distortion conditions, the correctness of all SNR levels within the same phase-changing condition were summed and then compared for statistical analysis.

3.5.3 Animal experiments and approach

Ethics statement

10 albino guinea pigs (200-800 g) of either sex were used in the experiments. Care and use of animals were carried out in accordance with the NIH Guide for the Care and Use of Laboratory Animals and were approved by the Animal Care and Use Committee of Northwestern University.

Surgery and electrode placement

Surgical procedures have been routinely performed in our laboratory[219-221]. Animals were anesthetized and the inferior colliculus of the guinea pigs was surgically accessed. A 16-channel thin film microelectrode (NeuroNexus, Ann Arbor, MI) was inserted for recordings into the central nucleus of the ICC perpendicular to its iso-frequency planes. A detailed description of the anesthesia and the animal monitoring is provided in the supplementary information.

Measurements

After a stable single unit in the ICC was identified, its BF was determined (detailed information seen in the supplementary information). Next, a sequence of 6 speech sentences was played to the left ear while the neural activity was recorded. S1 is the original speech sentence from the HINT test, "the silly boy is hiding". S2 to S6 were reconstructed based on the signal processing method described earlier. For each identified neural unit, the speech sentence sequence was repeated at least 100 times to achieve a sufficient number of action potentials for analysis. The timing of the action potential of individual ICC single units was sorted offline and transferred as neural pulses for analysis (see also data analysis below).

Data analysis

Firstly, neural coding of TFS was investigated. Sentences were divided into time sections based on the spectral properties to study the phase locking in ICC neurons. For example, a 200 ms time section containing "o" from "the silly boy is hiding" was selected for all six sentences (Figure 3-3). A Fourier transformation was performed on the selected speech sentence section to extract

the periodic patterns of TFS (Figure 3-1 and 3). The phase locking patterns to the period patterns of TFS could be derived from the spectral analysis of the neural PSTH. A Fourier transformation of the PSTH was evaluated. If dominating peaks were observed in the spectra (Figure 3-3(a), (c), (d)), the positions of these frequency peaks were taken to represent the frequencies at which neurons showed phase locking. SI was calculated to examine phase locking strength for each frequency of intests. A CC index could be calculated between the spectra of the acoustic signal and neural PSTHs. If no dominating peaks could be observed, phase locking was considered as absent.

3.6 Acknowledgments

3.7 Author contributions

C.P.R., Y.X., and M.C. designed the experiments, Y.X., M.C., X.T. and P.L. acquired data, Y.X., M.C. analyzed data, C.P.R., Y.X., M.C. and X.T. wrote manuscript, Y.X. and C.P.R. revised manuscript.

3.8 Supplementary Information

3.8.1 Hearing test for human experiments

To evaluate the hearing performance of subjects, their pure-tone hearing thresholds were determined, using an Audiometer GUI built in MATLAB. A sequence of two seconds pure tones

at 250, 500, 800, 1000, 2000, 4000, 8000, and 10000Hz was played to the test subjects. After a pure tone of each frequency was played, subjects were asked to respond whether they've heard the sound or not. The level of the tone was varied until the test person just identified the tone. Results were compared to a standard reference group verifying that no hearing loss exists. Subjects were tested with their left ear first and then right ear. Sounds played to both ears were the same in settings.

Devices and software used in this study were: Macbook Pro; MATLAB R2014b; headphone (Beyerdynamic DT990); sound level meter (Radio Shack) for course calibration and the Bruel& Kjaer 1/8 inch microphone for off-line recalibration of the headphones.

3.8.2 Animal experiments and approach

3.8.2.1 Surgery and electrode placement

All animals were anesthetized during experimental procedures. Anesthesia was induced by an intraperitoneal injection (i.p.) of urethane (0.9 mg/kg) in a 20 % Ringer's Lactate (RL) solution. During the experiment, the level of anesthesia was assessed by a paw withdraw reflex at 15 mins intervals. Supplemental doses of ketamine (40 mg/kg) and xylazine (2.5 mg/kg) were given with RL solution when needed. The body temperature of the animals was maintained at 38°C with a heating pad. A BM3-Vet system (Bionet Co. Ltd, Seoul, Korea) served to continuously monitor the animal's vital signs, including heart and respiratory rates and blood oxygen saturation.

The frontal bony skull of the animal was surgically exposed and mounted to a stereotactic head holder (Stoelting, Kiel, WI) using dental acrylic (Methyl methacrylate, Co-oral-ite Dental

MFG Co., CA). The inferior colliculus was surgically accessed through an opening on the right parietal bone just dorsal to the temporoparietal suture and just rostral to the tentorium. A 16-channel electrode (A1x16-5mm-100-177, NeuroNexus Technologies, Ann Arbor, MI) was used for single neuron recordings in the central nucleus of the inferior colliculus (ICC). The electrode was advanced into the ICC at 45° off the parasagittal plane using a 3D micromanipulator (Stoelting, Kiel, WI). After the initial placement, acoustic clicks at various levels were presented to the left ear. Neural activities were monitored on an oscilloscope to confirm whether the recording is from individual ICC neurons. The electrode was further advanced into the ICC in steps of 400 µm until a single neural unit was identified. The position of the electrode was maintained for further measurement.

3.8.2.2 Characteristic frequency (CF) assessments

After a single neuron had been identified, its CF was roughly estimated using sweep tones. Then pure tone ranging from 2 octaves above and below the estimated CF was presented to accurately assess the CF. The pure tone was given at levels ranging from 0 to 80 dB attention from the maximum speaker output (110 dB SPL) with 5 dB step. The number of evoked action potential was calculated to assess the CF.

3.8.3 Envelope coding remain comparable after phase shifting

Figure 3-S 1. PSTHs constructed from the neural activity recorded in the ICC while the sentences S1 to S6 were played to the ear.
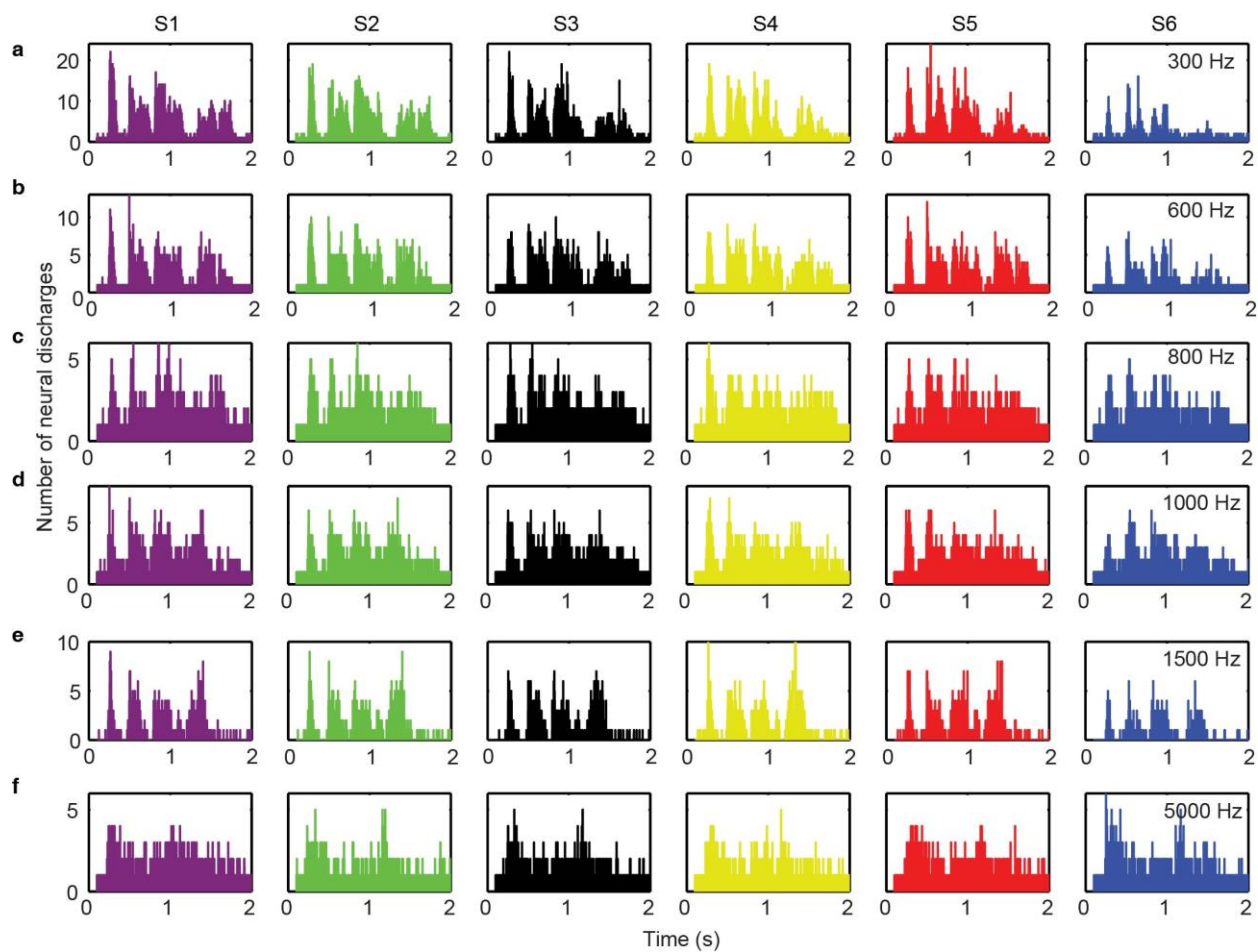
Figure 3-S1. Rows a-f show the results for six representative ICC neurons with different BFs.

Figure 3-S 2. The similarity of the neural PSTHs for S1 to S6 was compared using CC index across all individual neurons.
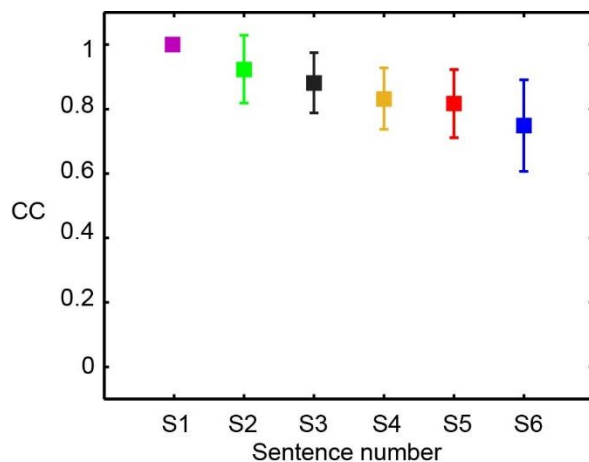
Figure 3-S2. The average CC index between neural PSTHs evoked by S1 and S1 to S6 was plottet with its standard deviation. The time window for PSTH was 10 ms.

3.8.4 The effects of the n-of-m strategy and phase manipulation on envelope, periodicity, and TFS within frequency bands:

To set the ground for discussion, a comparison of unfiltered acoustic signal and the filtered acoustic signal is shown in Figure 3-S3 to show the frequency range from 50 to 500 Hz. The phase distortion approach mainly targets the frequency range from 50 to 500 Hz, i.e., the periodicity defined in Rosen's 1992 paper. This terminology is also used by other studies (Rosen,1992; Langner, 1992; Green et al., 2005; Steinmetzger & Rosen, 2015; 2017a; 2017b). Since this range is most evident in low frequency bands with vowels, the voiced syllable "o" from "the silly boy is hiding" and a low frequency band (centered frequency of 689 Hz) was picked for demonstration.

The speech signal was filtered into 16 frequency bands up to 10 kHz. The band-pass filters were relatively broad resulting from the sound level used in the experiments. The sound level used

for human speech perception tests was 61 dB (re 20 µPa). The sound level used for neural activity recordings was 80 dB (re 20 µPa). These levels were relatively high, leading to a broad tuning for both human and guinea pigs. Previous studies that compared the sharpness of tuning between humans and guinea pigs reported inconclusive results. The broad tuning evoked in our study was aimed to overcome the possible difference of auditory tuning between the two species as addressed in the discussion of the manuscript.

The amplitude of the unfiltered signal and the filtered signal was plotted in its time waveform (Figure 3-S3 (a), (d)). The spectral analysis of this time waveform showed the amplitude fluctuation at different frequencies (Figure 3-S3 (b), (e)). The same neural response was plotted twice for the purpose of demonstration (Figure 3-S3 (c), (f)). The PSTH was constructed using a 0.025 ms time window, which is short enough to capture fluctuation below 20,000 Hz based on the Nyquist–Shannon sampling theorem. Spectral analysis of the PSTH demonstrated the phase locking patterns (Figure 3-S3 (c), (f)).

Figure 3-S 3. A voiced syllable "o" from "the silly boy is hiding" was selected (0.714 to 0.914 s post the onset of each sentence).
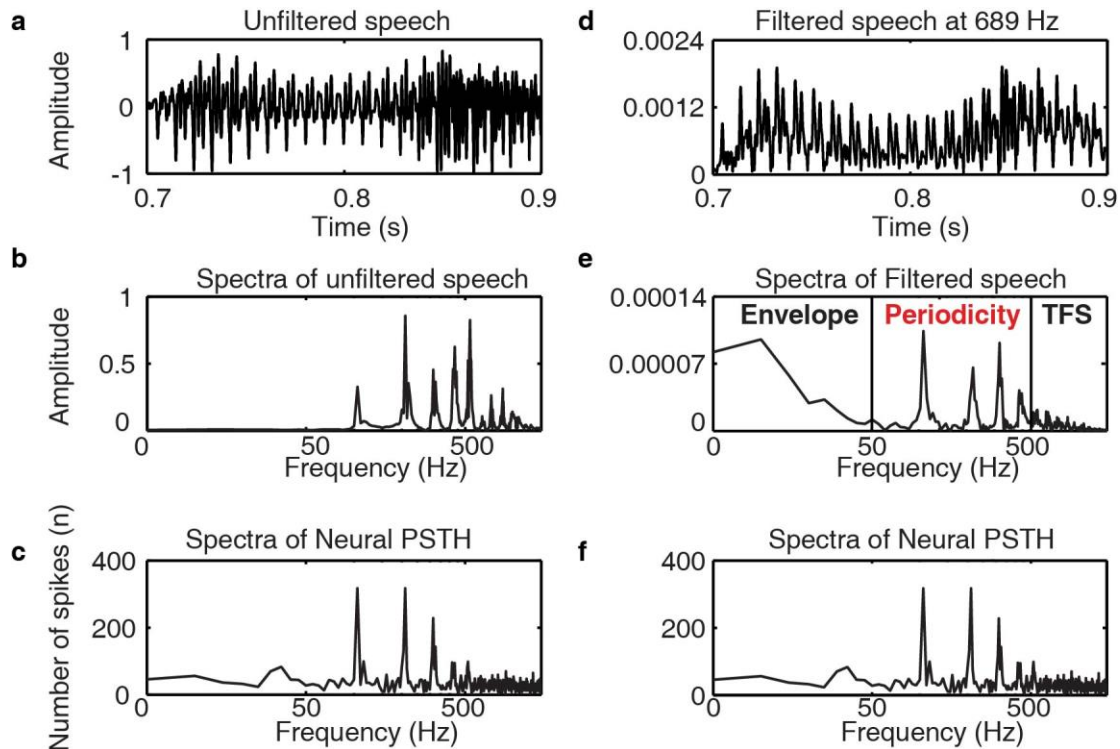
Figure 3-S3. a and d showed the time waveform of unfiltered speech and band-pass filtered speech at the center frequencies of 689 Hz. b and f showed the spectra of the unfiltered speech and band-pass filtered speech. The range of envelope, periodicity, and part of the TFS were listed. c and g showed the same spectra of a neural PSTH. The "periodicity" was referred as the periodic cues of TFS in the manuscript. The best frequency of the neuron was 600. The number of repetitions was more than 100 for the sentence.

As shown in Figure 3-S3(e), there are three components based on frequency ranges according to Rosen's paper (1992):

1. Envelope: The low frequency fluctuation <50 Hz. The envelope was better observed in the band-pass filtered signal. This is because the band-passing filtering at 689 Hz emphasized low frequency components more compared to the unfiltered signal.

2. Periodicity: The 50 to 500 Hz fluctuation, which is the target frequency range of our study. Within the range of 50 to 500 Hz, there were dominating frequency peaks in both unfiltered and filtered signals, i.e. the fundamental frequency and its harmonics. Due to the recognition of recent literatures refer TFS simply as the high frequency components of auditory filter outputs. The term "periodic cues of TFS" was used to substitute periodicity in the manuscript.

3. TFS: The high frequency range from 600 to 10,000 Hz, but was not fully plotted.

The effects of n-of-m strategy and phase distortion on the output of auditory filter at the center frequency of 689 Hz (Figure 3-S4) was presented. For the purpose of clarification, we use "envelope", "periodicity", and "TFS" as shown in Figure 3-S3.

The analysis of band-pass filtering on S1 to S6 was presented in Figure 3-S4. The amplitude of one frequency band (center frequency 689 Hz) was plotted in its time waveform for S1 to S6 (Figure 3-S4 (a) – (f)). The spectral analysis of the time waveform showed the frequency component of the amplitude fluctuation for S1 to S6 (Figure 3-S4 (g) – (i)). The spectral analysis of neural PSTH in response to S1 to S6 was shown (Figure 3-S4 (m) – (r)). As demonstrated by the spectral analysis, the envelope was nicely preserved from S1 to S2, S3, S4, but not in S5 and S6. The periodicity was preserved from S1 to S2, S3, but not in S4, S5, S6. TFS was slightly altered from S1 to S2-S6, but there was no obvious pattern. On the other hand, normal speech perception

was only in S1, S2, S3. Phase locking to fundamental frequency and its harmonics was only observed in response to S1, S2, S3. Thus, the speech perception results and phase locking patterns were greatly correlated with the change in periodicity, not with envelope. In other worlds, our approach has isolated the effects of the targeted frequency range 50 to 500 Hz.

The "discontinuities in the signal introduce the m-of-n strategy" could be seen through the comparison between S1 and S2 (Figure 3-S4 (a), (b), (g), (h)). The time waveform of original auditory filter output was shown in Figure 3-S4 (a). Discontinuities in the signal introduced by the m-of-n strategy resulted a time waveform shown in Figure 3-S4 (b). A better comparison can be seen in the spectral analysis (Figure 3-S4 (g) and (h)). The envelope (<50 Hz) and periodicity (50 to 500 Hz) were very comparable between S1 and S2. The TFS range was not changed dramatically. More importantly, the speech perception results for S1 and S2 are very similar. Thus, the slight alternation in TFS (> 500 Hz) didn't affect the speech perception results. Of note, the neural response and the phase locking patterns evoked by S1 and S2 were similar (Figure 3-S4 (m) and (n)).

Figure 3-S 4. A voiced syllable "o" from "the silly boy is hiding" was selected (0.714 to 0.914 s post the onset of each sentence) and filtered into 16 frequency bands up to 10 kHz.

a S1 Original speech — Amplitude vs Time (s), 7.7 to 7.9

b S2 n of m strategy — Amplitude vs Time (s), 9.7 to 9.9

c S3 Phase shift half period — Amplitude vs Time (s), 13.7 to 13.9

g Amplitude vs Frequency (Hz): Envelope | Periodicity | TFS

h Amplitude vs Frequency (Hz): Envelope | Periodicity | TFS

i Amplitude vs Frequency (Hz): Envelope | Periodicity | TFS

m Number of spikes (n) vs Frequency (Hz)

n Number of spikes (n) vs Frequency (Hz)

o Number of spikes (n) vs Frequency (Hz)

d S4 Phase shift one period — Amplitude vs Time (s), 11.7 to 11.9

e S5 Phase shift random — Amplitude vs Time (s), 15.7 to 15.9

f S6 Phase set to zero — Amplitude vs Time (s), 17.7 to 17.9

j Amplitude vs Frequency (Hz): Envelope | Periodicity | TFS

k Amplitude vs Frequency (Hz): Envelope | Periodicity | TFS

l Amplitude vs Frequency (Hz): Envelope | Periodicity | TFS

p Number of spikes (n) vs Frequency (Hz)

q Number of spikes (n) vs Frequency (Hz)

r Number of spikes (n) vs Frequency (Hz)

Figure 3-S4. One frequency band was selected as an example. Panels a to f show the time waveform of filtered S1 to S6. Panels g to l show the spectra of the filtered signal. Panels m to r show the spectra of the PSTHs constructed from the neural responses of one neuron. PSTH was recorded with 100 repartitions of the acoustic signal.

Chapter 4:

Neural representation of the temporal features of natural speech in the inferior colliculus

Yingyue Xu[1,3], Xiaodong Tan[1], Maxin Chen[2], Petrina LaFaire[1], Nan Xia[4], Claus-Peter Richter[1,2,3,*]

[1] Northwestern University, Department of Otolaryngology, 320 E. Superior Street, Searle 12-561, Chicago, IL 60611, USA.

[2] Northwestern University, Department of Biomedical Engineering, 2145 Sheridan Road, Tech E310, Evanston, IL 60208, USA.

[3] Northwestern University, The Hugh Knowles Center, Department of Communication Sciences and Disorders, 2240 Campus Drive, Evanston, IL 60208, USA.

[4] Institute for Digital Medicine and Computer-assisted Surgery, Qingdao University, Qingdao 266005, China.

This chapter is under preparation

4.1 Abstract

Natural speech, like many other biological significant sounds, contains harmonic signals that can be characterized by their fundamental frequencies and harmonics. These harmonic sounds evoke strong pitch sensations and play significant roles in speech perception in both quiet and noisy environments. Theories regarding pitch perception have suggested a transition from the predominant temporal coding at the auditory periphery to a rate coding at the higher centers. This transition has been suggested to occur in the central nucleus of the inferior colliculus (ICC), which integrates information from low level nuclei and projects to the thalamus and cortex. This indicates that temporal coding in the ICC differs from the auditory nerve. It has been shown that the neural activity of most ICC single units synchronizes with the fundamental frequency (F_0) when F_0 does not change. However, if F_0 changes, only ~47% of ICC units show synchronization to the F_0 of mandarin syllables and ~5% can follow the time-varying F_0. It remains unknown what's the proportion of ICC units which synchronizes to F_0 in response to natural English sentences with non-steady F_0. Our study systematically characterized the firing patterns of individual units from the ICC in response to natural English sentences. The data showed that 74% of ICC units showed synchronization to F_0. 48% ICC neurons can follow the time-varying F_0 in natural English speech. The low frequency speech envelope was coded by the average rate of neural discharges in up to 67% of the units of the ICC.

4.2 Introduction

Harmonic signals are common features(components) of human speech, music, and animal vocalizations [91, 113]. In speech, they convey semantic information and are critical for speech perception [114, 115]. Harmonic signals can usually be characterized by the combination of a fundamental frequency (F_0) and its integer multiples. The amplitudes and phases of these components contribute to the temporal features of the signal, which is always periodic at the frequency of F_0, even if F_0 is missing [133]. Harmonic signals can evoke strong pitch sensation at F_0. The neural coding of pitch sensation has been extensively studied at the level of individual auditory neurons in the cochlea [98, 99, 116-119], the cochlear nucleus [120-123], the inferior colliculus [124, 125] and the cortex [126]. The intensity of the pure tone or complex multi-tone sound is typically represented by the rate of action potentials and its frequency by the place along the cochlea, a principle named rate-place coding [129, 222, 223]. However, experiments have shown that firing rate and cochlear tonotopic organization are not sufficient to code harmonic signals. While cochlear responses show individual peaks in response to low-frequency resolved harmonic signals, responses to high-frequency unresolved signals cannot be separated. The neural coding of harmonic sounds also requires temporal coding, which can be identified by the synchronization of the neural activity of the ICC units through phase-locking and pattern in the inter-spike interval distribution [18-20]. Phase locking refers to the timing of neural discharges occurring at a certain phase of cyclical waveforms in response to periodic sounds.

Temporal coding can process both resolved and unresolved harmonic signals. Rate-place and temporal coding were observed in the auditory nerve [5-10] and the cochlear nucleus [11-14]. However, the predominant temporal coding in the periphery degrades and changes to rate-place coding in the cortex [17, 21]. It has been suggested that the majority of the units in the cortex encode only the lower frequency range relevant to periodicity pitch [22, 23]. It has also been suggested that cortical units processing slow modulations with rate-coding are restricted to a low-frequency region near the anterolateral border of the primary auditory cortex [17]. The ICC receives input from several peripheral auditory nuclei and relays information to the auditory cortex. It is likely the site where the transition occurred [24].

The neural temporal coding of harmonic signals has been widely examined with amplitude modulated sounds [113]. Phase-locked responses to low frequency modulation have been recorded in neurons located at the cochlea [98, 99, 116-119], cochlear nucleus [120-123], inferior colliculus [100], medial geniculate body [134, 135], and at various areas of the auditory cortex [136, 137]. The upper limit of phase locking is species dependent [140]. Auditory nerve fibers show robust phase locking to pure tone stimuli up to 9 kHz in barn owls [138], 4-5 kHz in cats [118], and 3.5 kHz in guinea pigs [99, 139]. Phase locking in auditory nerve fibers also correlates with the neurons' BFs: phase locking decreases for fibers with higher BFs [141, 142], and is maximal if the stimulus frequency is near the neurons' BFs [97].

The robust phase locking degrades in units in the ICC: 60-70% units show phase locking responses to pure tones with a cut off at about 1 kHz in guinea pigs [100]. Similar changes in synchronization for units of the ICC have been observed in other species [113]. Phase locking to modulations was seen in 30% of ICC neurons in anurans [143]. Langner and Schreiner's data in

cats showed that ~ 75% of the single units in ICC were tuned to a best modulation frequency when measured by their firing rate, in contrast to ~33% of the single units that were tuned to a best modulation frequency in their synchronization [124]. More recently, Su and Delgutte showed non-tonotopic rating coding in the ICC using harmonic complex sounds [133]. Their results indicate that the temporal coding is partially transformed into rate-coding at the level of ICC.

The neural temporal coding of harmonic signals has also been studied with speech tokens, such as synthesized vowels [99, 114, 139, 144, 145]. Auditory nerve fibers synchronize to F_0 and to high-frequency formants such as F1, F2, and F3 up to several kilohertz, as well as some of the distortion products including 2*F1, F1+F2, etc. [97, 101-105, 146, 147]. The strength of the phase lock depends on the amplitude of the harmonics and the neurons' BFs. In contrast to the auditory nerve fiber, the neural synchrony in the ICC is restricted to the low frequency harmonics of vowels. In response to a steady-state vowel /a/, most of the units in the cat ICC with a BF below 4 kHz are strongly phase locked to F_0 and its low harmonics (<1 kHz), but not to the higher formants [37] This confirms the significant loss of temporal coding at the midbrain. However, for BFs below 8 kHz, only ~47% of the guinea pig ICC units synchronize to F_0, and only ~5% of the ICC units phase lock to the time-varying F_0. This has been demonstrated in syllables of a tonal language (Mandarin, syllable /ba/ with four lexical tones: flat, rising, falling then rising and falling) [46].

At present, it remains unknown what proportion of the ICC single units phase locks to F_0 in response to natural English speech sentences, in which F_0 is non-steady but less variable than tonal language. In our study, we systematically examined the phase locking patterns of ICC units over a wide range of BFs with a natural English speech in guinea pigs. We also examined the neural representation of the slow varying speech envelope (E, frequency < 50 Hz [2]) through

neural firing rate, particularly identifying how each neuron synchronizes with a given integration period and range of frequencies.

4.3 Methods

4.3.1 Animals and surgeries

Ten albino guinea pigs (200-800 g) of either sex were used in the experiments. Care and use of animals were carried out in accordance with the NIH Guide for the Care and Use of Laboratory Animals and the procedures were approved by the Animal Care and Use Committee at Northwestern University.

Anesthesia was induced by an intraperitoneal injection (i.p.) of urethane (0.9 mg/kg) in a 20 % Ringer's Lactate (RL) solution. During the experiment, the level of anesthesia was assessed by a paw withdraw reflex at 15-minute intervals. Supplemental doses of ketamine (40 mg/kg) and xylazine (2.5 mg/kg) were given with RL solution when needed. The animal's body temperature was maintained at 38°C with a heating pad. A BM3-Vet system (Bionet Co. Ltd, Seoul, Korea) served to continuously monitor the animal's vital signs, including heart and respiratory rates and blood oxygen saturation.

Surgical procedures have been routinely performed in our lab [114, 219, 221, 224, 225]. The frontal bony skull of the animal was surgically exposed and mounted to a stereotactic head holder (Stoelting, Kiel, WI) using dental acrylic (Methyl methacrylate, Co-oral-ite Dental MFG Co., CA). The ICC was surgically accessed through an opening in the right parietal bone just dorsal to the temporoparietal suture and just rostral to the tentorium. Single neuron recordings in the ICC were completed with a 16-channel electrode (A1x16-5mm-100-177, NeuroNexus Technologies,

Ann Arbor, MI). The electrode was stepped into the ICC at 45° off the parasagittal plane using a 3D micromanipulator (Stoelting, Kiel, WI).

After the initial stereotactic electrode placement with the electrode tip just next to the surface of the occipital cortex, the electrode was then advanced through the cortex towards the ICC in steps of 400 µm until at least one single auditory unit was identified. Acoustic clicks were continuously delivered to the outer ear canal while the electrode's insertion depth was further advanced in smaller steps to maximize the response from the single unit(s). The optimized position of the electrode was then maintained for further measurement. After the set of measurements was completed, the electrode was retracted and reinserted into the ICC at a different site and the placement procedure was repeated and another set of measurements was captured. On average, measurements at 10 different sites were taken in one animal.

4.3.2 Recording procedures

After a single neuron was identified, its BF was estimated using pure tones sweeping from 100 Hz to 50 kHz. To accurately determine the BF, pure tones were presented at levels ranging from 30 to 100 dB SPL with a resolution of 5 dB. In addition to pure tones, a 2-second long natural speech sentence from the HINT test "the silly boy is hiding" was delivered to the ear. The sound level used for neural activity recording was 80 dB SPL. This sound level provides a broad tuning in the ICC units and many units with different BFs respond to the given stimulus and the spectrum of the vowels is stable [97]. Neural activity from ICC units was recorded with a 16-channel Plexon data acquisition system (Model MAP 2007-001, Plexon Inc., Dallas, TX) at 40,000 Hz/channel.

For each unit, the acoustic stimuli were presented 100 times. The action potentials were sorted offline (Offline Sorter, Plexon Inc., Dallas, TX) and the timing of the action potential was transferred as neural pulses for analysis.

4.3.3 Analysis

Speech sentences were processed to extract F_0 and its harmonics. First, the acoustic signals were divided into 64 frequency bands using a short-time Fourier transform (STFT) analysis. To achieve the best temporal resolution, the frameshift for the STFT was set initially to 1 bin, which corresponding to 0.023 ms at a sampling rate of 44.1 kHz. The frequency band with its center frequency similar to the BF of the ICC unit was then selected for further analysis. Next, a Fast Fourier Transformation (FFT) was applied to magnitudes of the selected frequency band and the corresponding periodic components were determined.

The phase locking patterns could be derived from the spectral analysis of the peri-stimulus time histograms (PSTH), which was constructed from the neural responses to the acoustic stimuli recorded from 100 stimulus presentations to the single units in the ICC. The smallest bin size for the histograms was 0.025 ms, which corresponds to the 40,000 Hz sampling rate of the recording system. The bin size can also serve as the temporal integration period for ICC neurons, which was increased from 2.5 to 25 ms in steps of 2.5 ms. An FFT was applied to the PSTH and the frequencies of the maxima in the resulting plots were determined. The frequencies of the maxima are the frequencies to which neurons phase lock. If no dominating peaks could be observed, phase locking was absent. The strength of phase locking was quantified by the synchronization index (SI), the ratio between the number of action potentials that synchronize to a certain frequency and

the total number of neural discharges [8, 14, 56]. The SI can be calculated from the PSTH spectra by dividing the magnitude of the spectra at the phase locking frequency from the magnitude at frequency equals to zero. A one-way analysis of variance (ANOVA) was calculated to examine whether the SI values were different across different BF ranges.

4.4 Results

4.1.1 Phase locking responses to speech in ICC units

A speech sentence from the HINT test was adopted as the acoustic stimulus. This sentence was "the silly boy's hiding" and was 2 seconds long. The time waveform of the speech sentences is shown in figure 1(a). The corresponding spectrogram was extracted by applying an STFT (figure 1(b)). The STFT separated the speech into 64 frequency bands from 0 to 22,050 Hz. The time-varying amplitude of each frequency band is shown in the corresponding spectrogram (figure 1(b)). The time-varying F_0 and its harmonics can be seen in the low frequency spectrogram of the speech signal (figure 1(c)). In response to the speech sentence, the neural response from individual ICC neurons was plotted as the PSTH from a hundred repetitions. A representative PSTH from an ICC neuron is shown in figure 1(d). The BF of this neuron was 600 Hz.

Figure 4- 1. An example of the speech signal and its corresponding neural response.
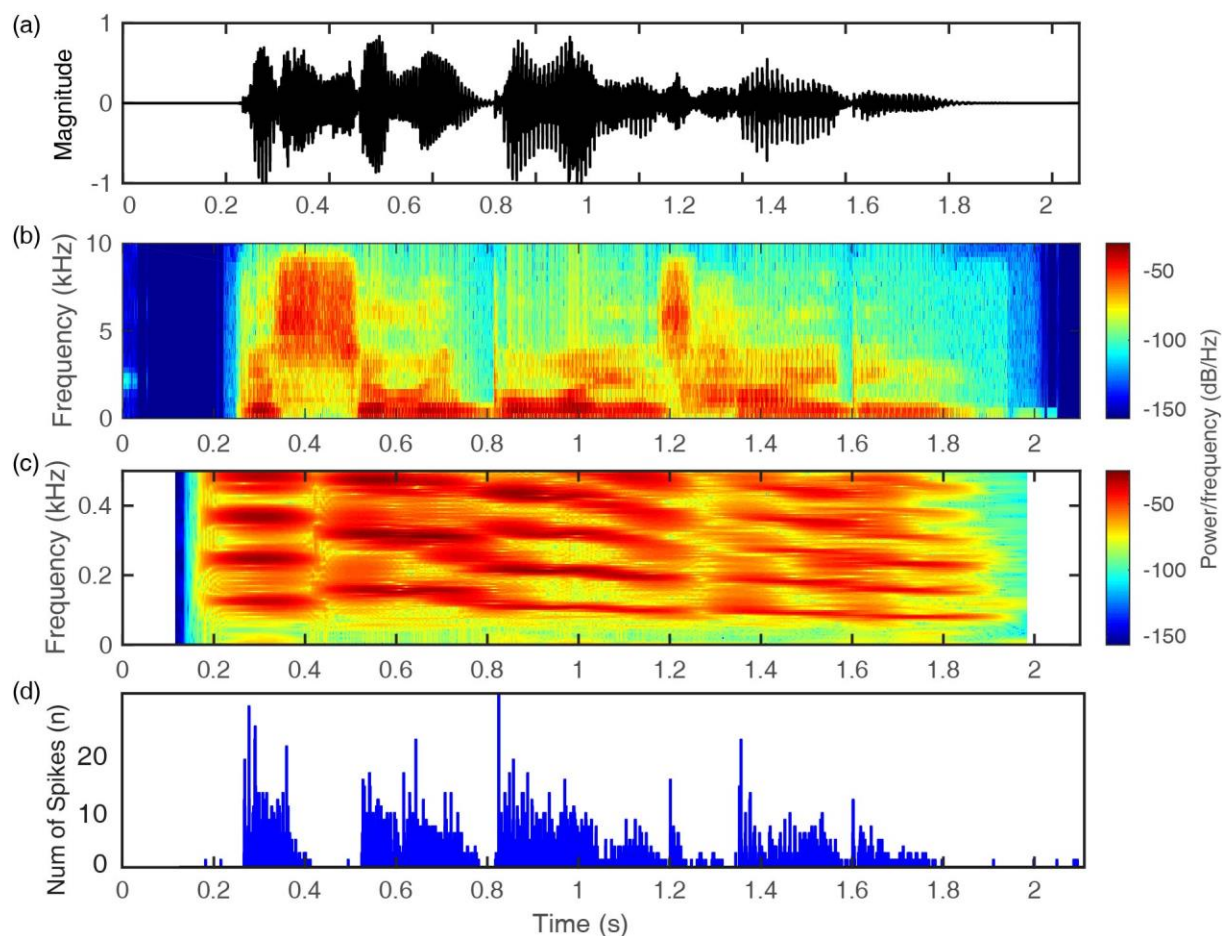
Figure 4- 1. (a) The time waveform and the content of the speech sentence "the silly boy's hiding" are shown. (b) The spectrogram of the speech signal. The spectrogram is plotted with time along the x-axis, frequency along the y-axis, and power/frequency shown as color. (c) The low frequency spectrogram of the speech signal (d) The peri-stimulus time histograms recorded from an individual ICC unit was plotted, showing the neural response to the given speech sentence. The BF of this unit was 600 Hz. The histogram was constructed from responses to100 stimulus presentations.

Single unit activity in response to pure tones has demonstrated phase locking patterns. An example from an ICC unit is given in figure 2. An FFT was performed on the neural PSTH to examine phase locking patterns. This unit, with the best frequency at 600 Hz, showed strong periodic firing at 600 Hz in response to the 600 Hz pure tone presented (figure 2(a)-(d)).

Figure 4- 2. ICC units demonstrated phase locking to pure tones.
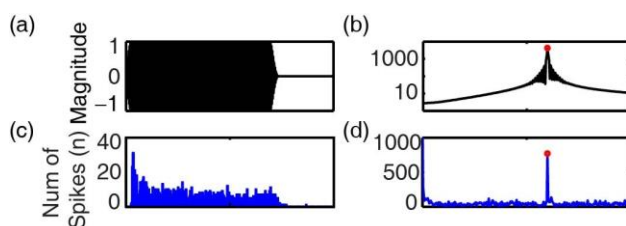
Figure 4- 2. (a) The time waveform of a pure tone at 600 Hz. (b) The spectra of the 600 Hz pure tone. (c) The neural response to the 600 Hz pure tone. (d) The spectra of neural PSTH in response to the 600 Hz pure tone.

The natural speech signal also evoked neural phase locking patterns. Due to the dynamic spectral transitions of speech, as shown in figure 1(b), speech phase locking was examined for different time sections based on the spectral properties. The voiced syllabus "oy" was selected as an example (figure 3(a)). This section was from 0.7 to 0.9 s in the 2-s long sentence and contained mainly low frequency information (figure 3(b)). The neural response from a representative unit is given as an example (figure 3(c)). T2he BF of this unit was 600 Hz. A corresponding frequency

band of the speech spectrogram was selected to match the BF of the unit (center frequency 689, closest to the unit's BF 600 Hz). The time-varying magnitude of this frequency band contains the slowly varying envelope (E, frequency < 50 Hz) as well as the fast varying temporal fine structure (TFS, frequency > 50 Hz) (figure 3(a)). The TFS features were periodic, and its dominating frequency components included the F_0 at 108 Hz and its harmonics (figure 3(b)). The neural PSTH evoked by the speech was also periodic (figure 3(c)). The spectrum of this PSTH shows a few dominating peaks that match the spectrum of the speech. In other words, the representative TFS frequencies were coded in the spectrum obtained from the PSTH of the neural activity (figure 3(d)), indicating phase locking to the F_0 and its harmonics. The strength of phase locking was quantified by the largest SI, i.e., the ratio between the magnitudes of the largest peaks of the PSTH spectra over the total number of action potentials. For example, the largest peak for this given unit was the second peak at 210 Hz. The magnitude of the second peak was 359, and the total number of firing for this time section was 718, giving the largest SI of 0.5.

Figure 4- 3. Phase locking response to speech signal.
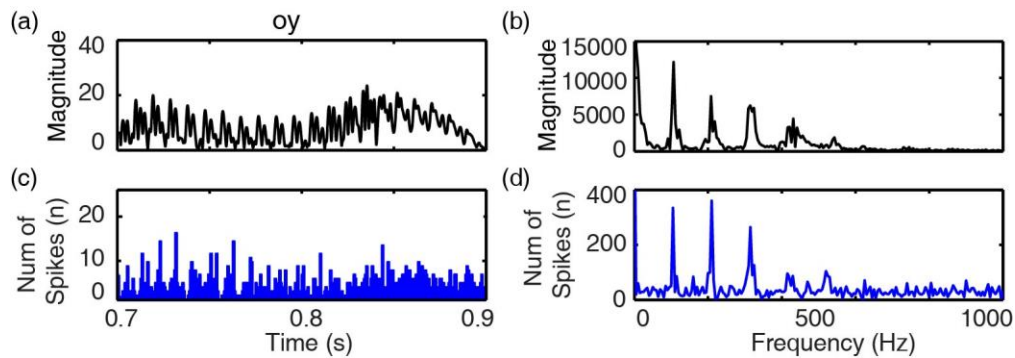
Figure 4- 3. (a) The time varying magnitude of a selected time section from 7.7 to 7.9 s, which corresponds to the "oy" from the sentence. (b) The spectra of "oy" showed characteristic frequencies of the TFS. (c) The neural PSTH to this stimulus. (d)The spectra of the neural PSTH showed dominating frequency peaks that matched the spectra of the speech, indicating phase locking to these characteristic frequencies of the TFS. The BF of this unit is 600 Hz. The histogram was constructed from responses to100 stimulus presentations.

Phase locking was observed in units with various BFs. Simultaneous recordings from several individual units with various BFs were achieved by using the 16-channel electrode. For example, three single units were identified during one set of recordings. Their BFs were 600, 1500, and 2800 Hz, respectively. Based on their BFs, the corresponding frequency bands were extracted from the speech signal. The spectra of each frequency band contained F_0 and its harmonics (Figure 4(a)-(c)). These dominating frequencies of the speech were coded in the corresponding units (Figure 4(d)-(f)). Of note, the absolute magnitude of F_0 and its harmonics dropped with an increase of the center frequency, the magnitude of neural PSTH didn't follow this trend. The absolute magnitude of neural PSTH depends more likely on the firing rate and synchrony of units.

Figure 4- 4. Phase locking response to speech signal recorded from multiple neurons.
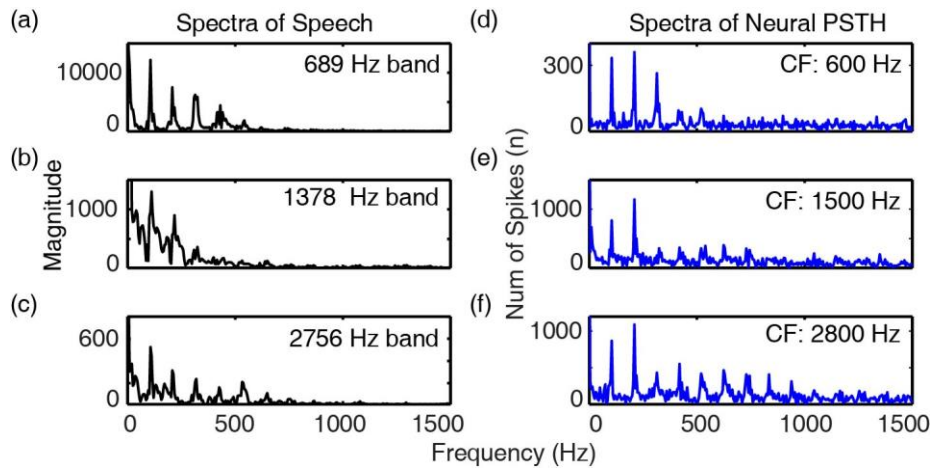
Figure 4- 4. (a), (b), and (c) The spectra of the TFS for three different frequency bands with center frequencies of 689, 1378 and 2756 Hz, respectively. (d), (e), and (f) show the spectra of neural PSTH of 3 units with the best frequency of 600, 1500, 2800 Hz. The histogram was constructed from responses to100 stimulus presentations.

Phase locking was examined in all 93 individual ICC units tested. The peak SI was calculated from three individual frequency ranges, 34 to 145 Hz, 145 to 245 Hz, and 245 to 345 Hz, to capture the F_0, the 2nd or 3rd harmonics of F_0. The peak SI of these three frequency ranges was plotted with their corresponding frequencies for all units tested (Figure 5(a)). The majority of the units showed peaks around the anticipated F_0, the 2nd or 3rd harmonics of F_0. The outliers indicate that no visible peak was observed at the targeted frequency ranges. It was likely artifacts due to spikes in the noise floor. The percentage of the single units that showed visible peaks at targeted frequency range was 74%, 66%, and 67% for F_0, the 2nd and 3rd harmonics of F_0 (Figure 5(b)). The largest SI values among these three peaks were shown in Figure 5(c) for all units tested.

The averaged maximum SI was 0.24±0.16. Phase locking was observed in units with different BFs (Table 1). The majority of ICC units have BFs below 5 kHz. The proportion of single units that showed phase locking to F_0 was relatively similar across different BF range from 0 to 20 kHz. One-way ANOVA test showed no significant difference for the SI values across different BF ranges (the f-ratio is 0.55 and the p-value is 0.65).

Figure 4- 5. The phase locking pattern to the speech of all units tested.
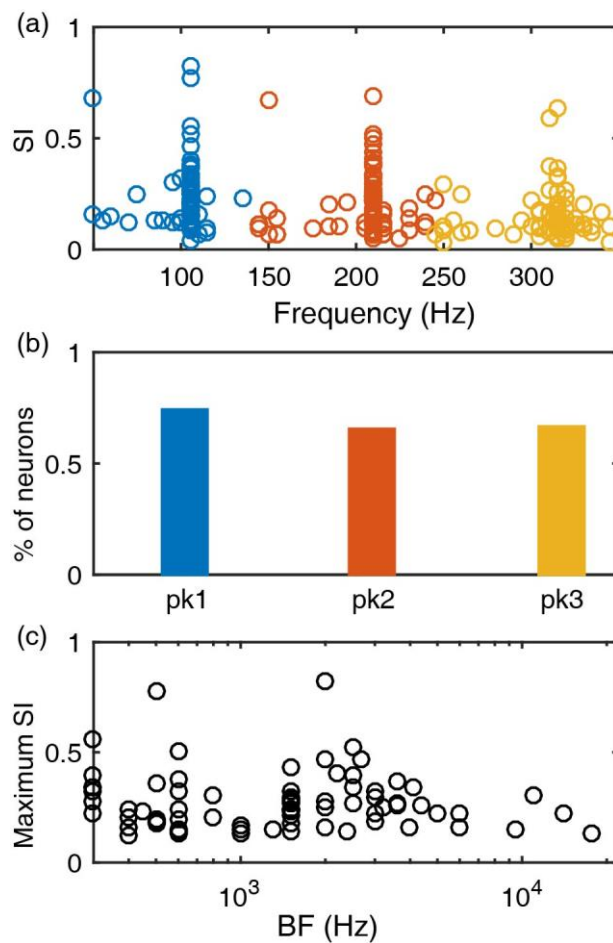
Figure 4- 5. (a) The largest SI within three individual frequency ranges: 34 to 145 Hz, 145 to 245 Hz, and 245 to 345 Hz. (b) Through a visual examination of the neural PSTH spectra, the percentage of units that showed a visible peak around F_0, the 2nd or 3rd harmonics of F_0 was plotted. (c) The largest SI among the visible peaks within three examined frequency ranges was plotted for each unit with their corresponding BFs. Units with no visible peaks in the PSTH spectra (26% of the tested ICC units) were not included.

BF range (kHz)	No. of single units phase locked to F0 / total units	Proportion of units	SI
0 – 1.25	30 / 34	88%	0.25 ± 0.16
1.25 – 2.5	25 / 34	74%	0.24 ± 0.15
2.5 – 5	13 /16	81%	0.20 ± 0.10
5 – 20	6 / 9	67%	0.20 ± 0.06

Table 4- 1. The summary of ICC single units tested: BFs, number of single units that showed phase locking to F_0, the total number of single units, the proportion of single units that phase locked to F_0, as well as the corresponding average SI and their standard deviations.

To assess the temporal coding of the time-varying F_0 in natural speech stimuli, the low-frequency spectrograms of neural PSTH was calculated. Two ICC units with the same BFs were shown as representative examples (figure 6). The spectrogram of neural PSTH resembled the

speech spectrogram for one ICC units within the time frame of 0.6 to 1.7 s (figure 6(b)), indicating phase locking response to the time-varying F_0. Such similarities between the speech and neural PSTH were observed in other units that showed phase locking. Temporal coding of the F_0 was not observed in some single units (figure 6(c)).

To quantify to which degree the ICC units follow the F_0, a normalized 2-D cross-correlation (CC) was calculated between the spectrogram of speech and neural PSTH. The results of normalized 2-D CC were plotted in figure 6(d) and (e). A local peak (figure 2(d)) can be seen for ICC units that followed F_0 (figure 6(b)), while no peak (figure 6(e)) can be seen in ICC unit that showed no phase locking to F_0 (figure 6(c)). There were only 7 ICC units that yielded a CC above 0.2 and 21 ICC units yielded a CC above 0.1. Although the CC values were low, 45 ICC units showed local CC maxima that resembled figure 6(d) based on visual examination, whereas the other 48 showed no obvious peaks.

Figure 4- 6. The temporal coding of time-varying F_0 in ICC units.
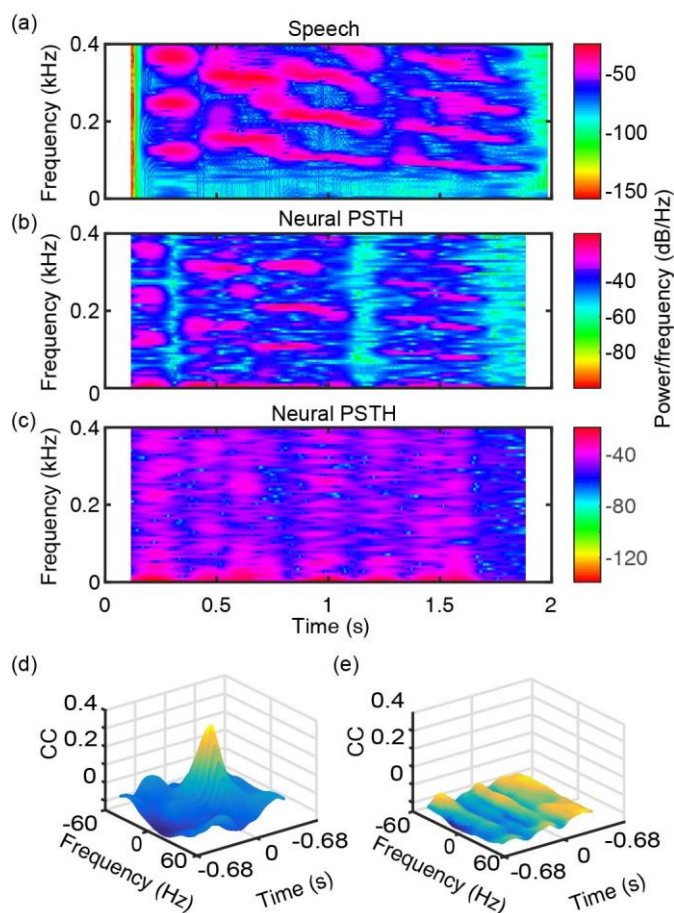
Figure 4- 6. (a) The low frequency spectrogram of the speech signal. The spectrogram is plotted with time along the x-axis, frequency along the y-axis, and power/frequency shown as color. (b) The PSTH of an individual ICC unit in response to the given speech sentence. (c) The PSTH of an individual ICC unit in response to the given speech sentence. The BF of the two units were 1500 Hz. The two histograms was constructed from responses to 100 stimulus presentations. (d) The CC between the spectrogram of the speech and neural PSTH shown in subpanel (b). (d) The CC between the spectrogram of the speech and neural PSTH shown in subpanel (c). The CC indexes for all the 93 ICC units were also compared with their BFs (figure 7(a)) and maximum SIs

(figure 7(a)). The CC values were relatively larger in low BF units compared with high BF units.

The CC values also increased with the increase of maximum SI.

Figure 4- 7. The results of normalized 2-D cross-correlation plotted as a function of ICC units' BFs (a) and maximum SI (b).
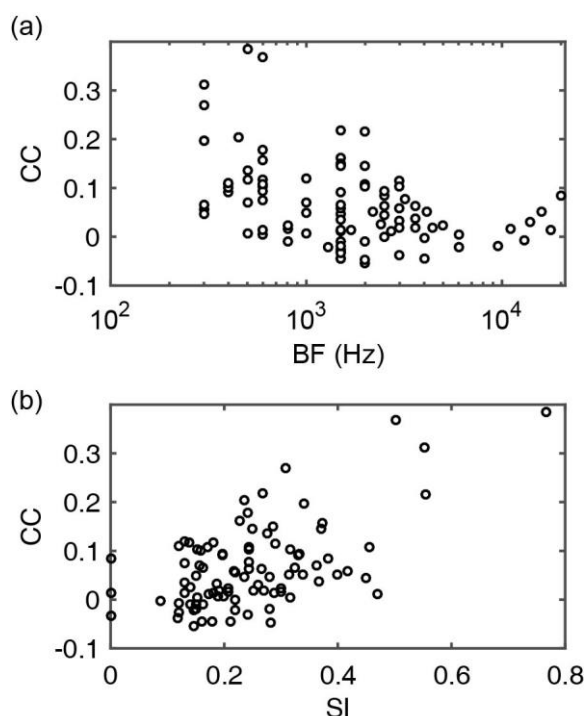
4.2.2 Neural firing rate encodes speech envelope in ICC units

To evaluate the speech E coding in ICC units, a detailed comparison of the neural firing rate and speech E was carried out. The original signal (Figure 8(a)) was filtered with a series of acoustic bandpass filters with 9 different bandwidths (0.1, 0.125, 0.33, 0.5, 1, 2, 3, 4, 5 octaves). These bandpass filters were Butterworth filters with a filter slope of -30 dB per octave [226]. The center frequency of the filter was matched with each ICC neuron's BF to best approximate the auditory filter. A representative unit with BF of 600 Hz was selected as an example, and the

bandpass filters were centered at 600 Hz. The lowpass cutoff the acoustic bands did not exceed 22,050 Hz due to the sampling rate of the speech signal. The filtered signal is shown in Figure 8(b). The bandpass filtered signal was then processed with a Hilbert transformation to extract the magnitude of the signal (Figure 8(c)). Furthermore, a lowpass filter with the cutoff frequency at 50 Hz was applied to extract the slow varying speech E (Figure 8(d)). The neural PSTH was extracted from the single unit response to characterize the rate of neural firing (Figure 8(g)). The PSTH bin width serves as the temporal integration period. The PSTH was processed with 10 bin sizes ranging from 2.5 to 25 ms with a step of 2.5 ms. The spike count increased with the increase of the PSTH bin size. A visual similarity between speech E and neural PSTH could be observed (Figure 8(d), (e)), which was quantified by Pearson correlation coefficient (CC). CC was calculated between the neural PSTH and the speech E. Note that the lag of PSTH relative to the acoustic stimuli was not corrected when calculating CC, which might result in slightly reduced CC values. The combination of nine acoustic bandwidths (Figure 8(d)) and ten PSTHs bin sizes (Figure 8(e)) yielded 90 combinations for CC calculations, as shown in Figure 8(f). These correlations were above 0.5 for this given ICC unit, showing a strong similarity between speech E and neural PSTH. The correlation calculation was performed on all ICC units. 62 out of the 93 tested ICC units showed at least one correlation larger than 0.5, indicating the majority ICC units could encode the temporal cue E through the neural firing rate.

Figure 4- 8. A detailed comparison between the speech envelope and neural peri-stimulus time histograms was performed.
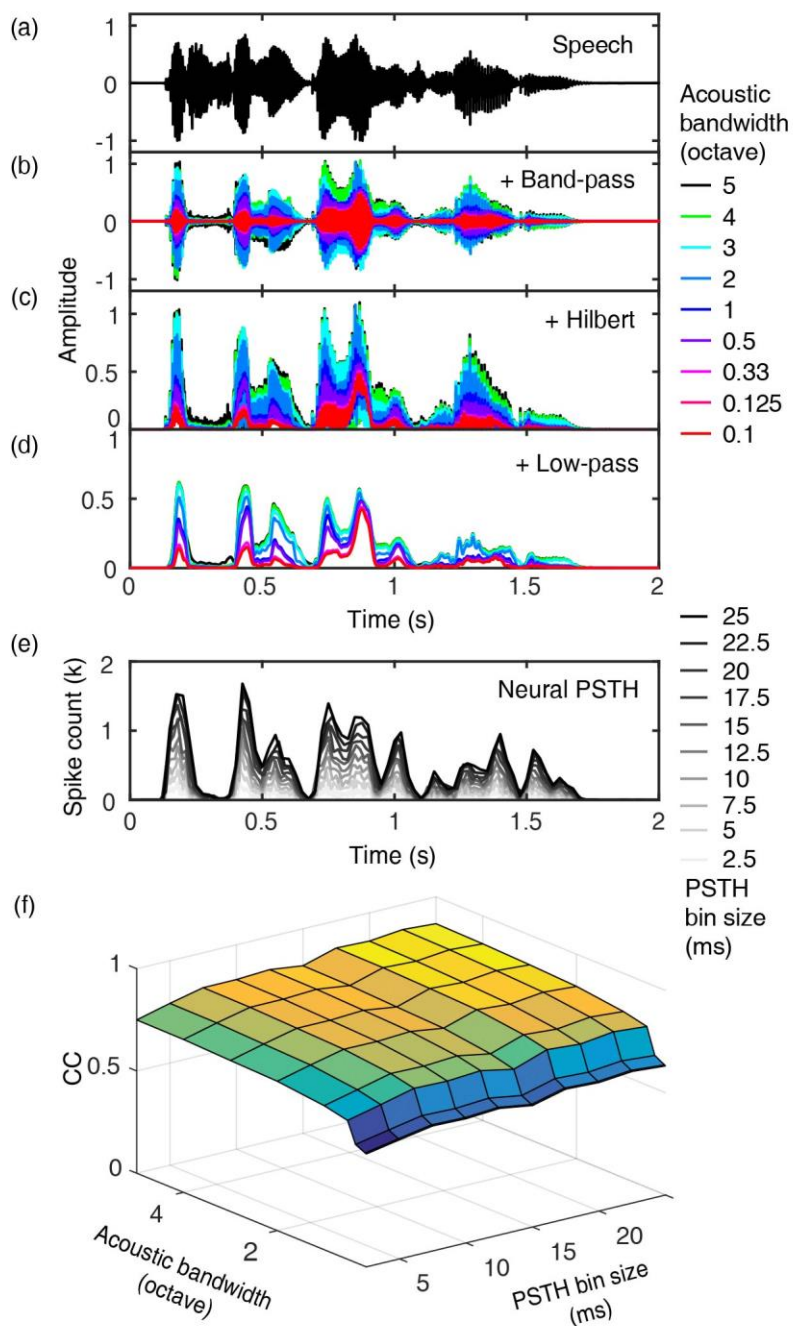
Figure 4- 8. The same color coding for bandwidth is used for the panel (b) to (d). (a) The

original speech signal is plotted versus time. (b) The speech signal was filtered with nine series of

bandpass filters of different bandwidths. The bandpass filters' center frequencies were set as 600

Hz. (c) Filtered speech signals were further processed with a Hilbert transformation. (d) A lowpass filter was applied after the Hilbert Transformation, which resulted in the slow varying speech E. (e) The peri-stimulus time histograms of a representative unit were plotted with various histogram bin sizes ranging from 2.5 to 25 ms. The best frequency of this unit was 600 Hz. The histogram was constructed from responses to100 stimulus presentations. (f) Pearson correlation coefficient calculated between the PSTH and the speech E was plotted as a surface plot for each of the 90 combinations of the ten PSTH bin sizes and the nine bandwidths of the acoustic filters. The color coding matches the correlation coefficient values.

The value of the correlation varied with PSTH bin size, as well as acoustic bandwidths. The correlation increased steeply with the increase of acoustic bandwidth from 0.1 to 0.5 octaves. Note that the CC values for 0.1 and 0.125 octaves are overlapping (Figure 8(f)). The correlation increased gradually from 0.5 to 5 octaves. This indicated that a good approximate towards the auditory filter could be achieved with 0.5-octave bandpass acoustic filter (30 dB per octave). Acoustic information beyond the 0.5-octave band might not be contributing to the auditory filter output of this unit. The correlation also increased with the PSTH bin size. A steep increase of CC could be observed with the bin size of PSTH from 2.5 to 5 ms. The CC increased gradually with the increase of PSTH bin size after 5 ms.

Furthermore, the minimum PSTH bin size and acoustic bandwidth that generated a large correlation (CC > 0.5) were examined in the 62 neurons that showed such large correlation. There were 61% of the ICC neurons showed that with a small PSTH bin size of 2.5 ms, the correlation above 0.5 could be achieved (Figure 9(a)). This was consistent with the previous hypothesis that

the central auditory system temporally integrates sound information within a window of roughly 5 ms [197, 227-229]. The temporal integration time window varied with the ICC neurons' BFs. The minimum PSTH bin sizes were small for ICC neurons with BFs lower than 1500 Hz but showed large variation beyond BFs above 1500 Hz (Figure 9(b)). 63% of the ICC neurons showed a correlation above 0.5 with an acoustic filter within the 1-octave band. Acoustic information beyond the 1-octave band might not be contributing to the auditory filter output of ICC neurons. The minimum acoustic bandwidths that were capable of approximating the auditory filter also varied with BFs of ICC neurons. The minimum acoustic bandwidths were small for ICC neurons with BFs lower than 1000 Hz but showed large variation with BFs above 1000 Hz (Figure 9(d)).

Figure 4- 9. Analysis of bin size and acoustic bandwidth.
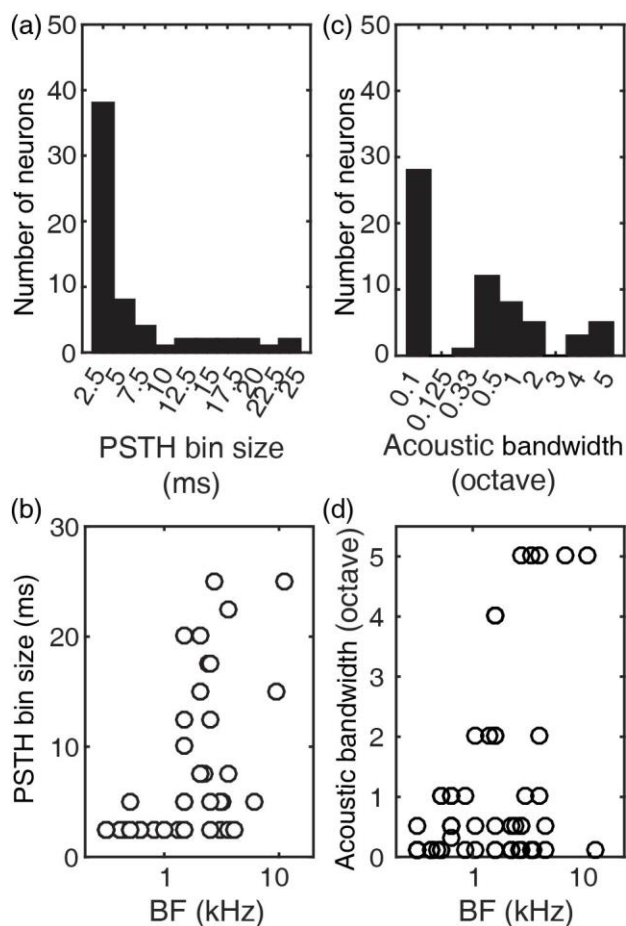
Figure 4- 9. (a) The histogram of the minimum bin size of the neural peri-stimulus time histograms that yielded a CC larger than 0.5 was plotted. 38 out of the 62 neurons yielded a large CC with PSTH bin size of 2.5 ms. (b) The minimum PSTH bin size that yielded a CC larger than 0.5 was plotted against neuron's BFs. (c) The histogram of the minimum acoustic bandwidth that yielded a CC larger than 0.5 was plotted. 28 out of the 62 neurons yielded a large CC with a 0.1-octave band. (d) The minimum acoustic bandwidth that yielded a CC larger than 0.5 was plotted against neuron's BFs.

4.5 Discussion

Through an examination of neural firing rate as well as speech E, a large correlation between the PSTH and speech E was observed. This further provided support that the ICC neurons can still synchronize to the envelope of natural speech signals. The analyzed PSTH bin size served as different arbitrary neural integration periods. The minimum integration period that generated a large CC (>0.5) was mainly 2.5 to 7.5 ms (Figure 9(a)), which is consistent with previous findings that the temporal integration period for central auditory neurons is approximately within 5 ms [197, 227-229]. Moreover, the integration time was small for ICC neurons with BF lower than 1,500 Hz (Figure 9(b)) and was variable for high frequency ICC neurons. Previous studies discussed the possibilities that some of the temporal cues might be coded spatially along the iso-frequency ICC laminae, such as periodicity and latencies under certain stimuli [124, 230-233]. The differences in temporal coding within each ICC lamina might be underlying the large variation in neural integration time across the high frequency ICC neurons.

The neural representation of speech E was also evaluated at various acoustic bandwidths. An acoustic bandwidth within 1-octave could generate a large CC (> 0.5) for the majority of the neurons (Figure 9(c)). Also, the minimal acoustic bandwidths that gave a large CC (> 0.5) was small for ICC neurons with BF lower than 1,500 Hz (Figure 9(d)) and was variable for high frequency ICC neurons. Of note, the sound level used in our experiment was at 80 dB SPL, which might widely excite an auditory filter. Nevertheless, this observation indicates that ICC neurons could encode speech E that derived from an acoustic band within 1 octave at this stimulating sound level.

Phase locking patterns have been intensively studied in the cochlea to uncover temporal processing of the auditory system. The stimuli for those experiments ranged from pure tones to phonemes and syllables [86, 97, 106, 116, 117, 123, 234-239]. Auditory nerve fibers' phase locking patterns depended on both the stimulus frequencies and the neurons' properties [99, 101, 240-245]. For example, synthesized vowels can evoke phase locking to the primary frequencies (fundamental frequency and formants). Additionally, phase locking was also observed in the nonlinear combination of these primary frequencies, which was suggested to originate from the nonlinearity of the cochlea. Furthermore, phase locking to certain frequencies could be enhanced if such frequencies are close to the neuron's BF. On the other hand, our data from the midbrain showed a different phase locking pattern. Individual ICC neurons showed phase locking to mainly the F_0, the 2nd or 3rd harmonics of F_0. Phase locking to frequencies higher than 1 kHz was not observed for speech, which might explain the lack of phase locking to high frequency formants and distortions observed in the auditory nerve fibers. Previous studies showed that phase locking to the periodic cues in speech is critical for speech perception in quiet and noisy listening conditions [114].

Phase locking to the F_0 of speech was found in 74% of ICC neurons recorded. This resembled previously reported proportions of ICC neurons that showed phase locking to pure tone stimuli [100]. In response to natural speech sentences, the ICC neurons showed phase locking mainly to the F0 and its 2nd and 3rd harmonics. Phase locking was also observed in the higher order of harmonics (Figure 4(f)) but never exceeded 1000 Hz. This observation is independent of the BFs of the neurons. Note, the sound level of the speech was 80 dB SPL, high enough to stimulate ICC neurons with low frequency speech cues. The 1 kHz frequency limit for phase

locking for units in the guinea pig ICC and the finding that the ability for phase lock is independent of the units' BFs were consistent with previous observations of pure tones in the ICC of guinea pigs [100]. The value of the SI provided the strength of the phase lock in response to the speech stimuli (Figure 5(c)) and was for most units below 0.5. This is different from the SI reported for pure tone stimuli, which was larger than 0.5 in many neurons. The smaller SI values for speech stimuli were likely due to neurons phase locking to multiple frequencies in the signal, reducing its ability to synchronize to each of the frequencies in the speech signal. Such observation could also be seen from the comparison of pure tone and multiple tone stimuli (Figure 2(d), (h)). This phenomenon also suggested that for complex acoustic stimuli, it is beneficial to have a relatively large number of action potentials to examine phase locking to multiple frequencies.

4.6 Conclusion

In summary, our study examined neural coding of natural speech temporal features in the auditory midbrain. Our data suggested that 67% of the ICC neurons encode the E of speech through its firing rate. These neurons code the speech E mainly with a neural integration period of 2.5 ms with a large variation for neurons with high BFs. Our data also suggested that up to 74% of the ICC neurons synchronize to temporal fine structure of speech. These neurons encode mainly the periodic components of speech TFS through phase locking to the F_0 and its 2nd and 3rd harmonics.

4.7 Acknowledgments

Chapter 5: Conclusion

Towards the development of an optical cochlear implant, there are two challenges that have been addressed in the dissertation: firstly, the design, fabrication and testing of a light delivery system with small optical sources that can be implanted into the cochlea and evoked auditory responses; secondly, the development of a coding strategy with temporal periodic cues to accommodate for the low rate of INS. We have also demonstrated that the temporal periodicity is essential for speech perception in normal hearing listeners in both quiet and noisy environment. In this chapter, we discuss the results shown this dissertation and provide directions for future work.

5.1 The development of a light delivery system

With our work, we have demonstrated that a light delivery system can be fabricated for optical stimulation for INS or optogenetics. The electrodes are small enough to fit into a human cochlea. The experiments have also demonstrated some challenges in using small optical sources for stimulation.

5.1.1 Design the light delivery system

In chapter 2, we have shown how to fabrication a circular optrode not exceeding the dimensions of scala tympani of the human cochlea [77, 179-181]. The most recent prototype is a 15 channel optrode with dimensions similar to commercially available CI electrodes (Figure 5-1).

The dimensions of the optrode should taper from 1 mm to 0.47 mm. The length of the electrode/optrode should be shorter than 27 mm. This measurement served as the upper limit of the optrode. Of note, an optrode design that is too small would bring difficulties during the insertion because the tip of the optrode would curl in the base of scala tympani. Continuing efforts are to align the physical design of the optrode with the dimensions of scala tympani of the human cochlea. We have matrix printed 3-d mold based on a human and cat scala tympani for optrode fabrication. The physical measurement and mechanical properties are currently being tested.

Figure 5- 1. shows a comparison between the tips of an optrode made with 15 light sources (top) and two commercially available electrodes.

Future studies on LDS design can focus in the following three directions: 1) physical design, 2) waveguides, and 3) hybrid array with electrical and optical stimulators. Firstly, the physical design of the LDS could be tailored to each patient in the future. An exemplary concept is given

here: prior to the cochlear implant surgery, a micro-CT scan could be taken to capture the 3-d image of the patient's cochlea. Then, the person's individual cochlear implant electrode would be created with a matrix printer to provide perfect fitting. This would allow the optimal placement and orientation of the optical sources towards the auditory neurons. Secondly, in addition to the LDS with small light sources, waveguides could be an alternative candidate towards optical cochlear implants. Waveguides are compliant for deep insertion into the cochlea. The challenge lies in coupling losses of the light source with individual waveguides. Thirdly, future LDS might also include a hybrid tip which includes both electrical and optical stimulators. The combined optical and electrical stimulation can lead to sub-threshold stimulation [184, 185].

5.1.2 Energy requirements and opto-electrical hybrid electrode.

In a recent paper, we reported the radiant energy required for INS in guinea pigs and cats [77]: between 4.1 ± 1.9 to 14.1 ± 8.1 µJ/pulse for ICC single units and 7.2 ± 4.7 to 17.2 ± 13.9 µJ/pulse for CAP responses depending on the distance between the light sources to target. As reported in Chapter 2, the radiant energy of the SELDs emitting light at 1850 nm was typically in the range between 8 and 20 µJ/pulse, which can reach the INS thresholds and is below the threshold for which cochlear damage was detected [182]. Additional efforts are needed to reduce the amount of energy required for stimulation. There are currently a few approaches that are under testing in the group. One possible approach is combined optical and electrical stimulation, which can lead to sub-threshold electrical stimulation [184, 185]. We have also shown in deaf white cats that combined optical and electrical stimulation reduces the threshold for INS in the cochlea [130].

Another approach to reduce the radiant energy for INS is to modify the pulse shape of the optical pulses. Experiments in our lab compared square with triangular pulses, indicating that triangular pulses evoked 5-times larger compound action potential amplitudes at the same radiant energy. Additionally, wavelength of the INS could be optimized to reduce the threshold. INS with wavelengths around 1860 nm has been commonly used in the thesis and existing literature due to its penetration depth in water. Other wavelengths in the infrared range that share similar penetration depth are around the wavelength for communication, i.e., wavelengths ranging from 1260 to 1675 nm. The technology is more mature for these wavelengths, yielding higher wall-plug efficiency. Thus, infrared laser with a wavelength between 1260 to 1675 nm might serve as a superior alternative compared with wavelength at 1860 nm. We have explored three wavelengths in a recent study: 1375 nm, 1460 nm, and 1550 nm. Our results have shown that the efficiency of INS is the highest at 1375 nm compared with 1460, 1550, and 1860 nm. Future studies will focus on adopting this wavelength for the development of optical cochlear implant electrodes.

5.2 Examine the function of speech temporal periodicity and its neural processing

5.2.1 The neural response from individual auditory neurons can be used to analyze neural processing to the periodic cues of the TFS

The neural temporal coding of is usually assessed in two measures: the average discharge rate (i.e., the average number of neural discharges over modulation periods), and neural

synchronization or phase locking. Phase locking refers to the timing of neural discharges occurring at a certain phase of cyclical waveforms in response to periodic sounds. The neural response from individual auditory neurons can reflect both "neural synchrony to the envelope structure" and "neural synchrony to periodic cues of the TFS". E and TFS are essentially low and high frequency components of an acoustic signal within the auditory filters. Both signals are time varying sound pressure levels in the time domain. Based on the Nyquist–Shannon sampling theorem, E (< 50 Hz) can be represented by a low sampling rate (eg. 100 Hz), whereas the TFS requires a high sampling rate for faithful reproductions. The neural discharges in the auditory neurons could be viewed as a neural sampling of acoustic signals. The (peri-stimulus time histogram) PSTH serves as the neural representation of the acoustic signal. The bin size of the PSTH gives the time resolution of the PSTH, namely the sampling rate of the neural representation of the acoustic signal. For example, a PSTH with a bin size of 10 ms gives a 100 Hz sampling rate, which can encode a frequency range of 0 to 50 Hz. An example has been given in Figure 5-2. The black line (Figure 5-2(a)) shows the acoustic signal in the time domain. Neural PSTH with a 10 ms bin size is plotted in Figure 5-2(b). This PSTH has a low time resolution and reveals the "neural synchrony to envelope structure". On the other hand, a PSTH with a small bin size would reveal not only the "neural synchrony to envelope structure", but also the "neural synchrony to periodic cues of the TFS. The example is given in Figure 5-2(c). This PSTH has a bin size of 0.25 ms, which give a 4,000 Hz sampling rate. This sampling rate can encode frequency range up to 2,000 Hz based on the Nyquist–Shannon sampling theorem.

Figure 5- 2. Speech signal and neural PSTH.
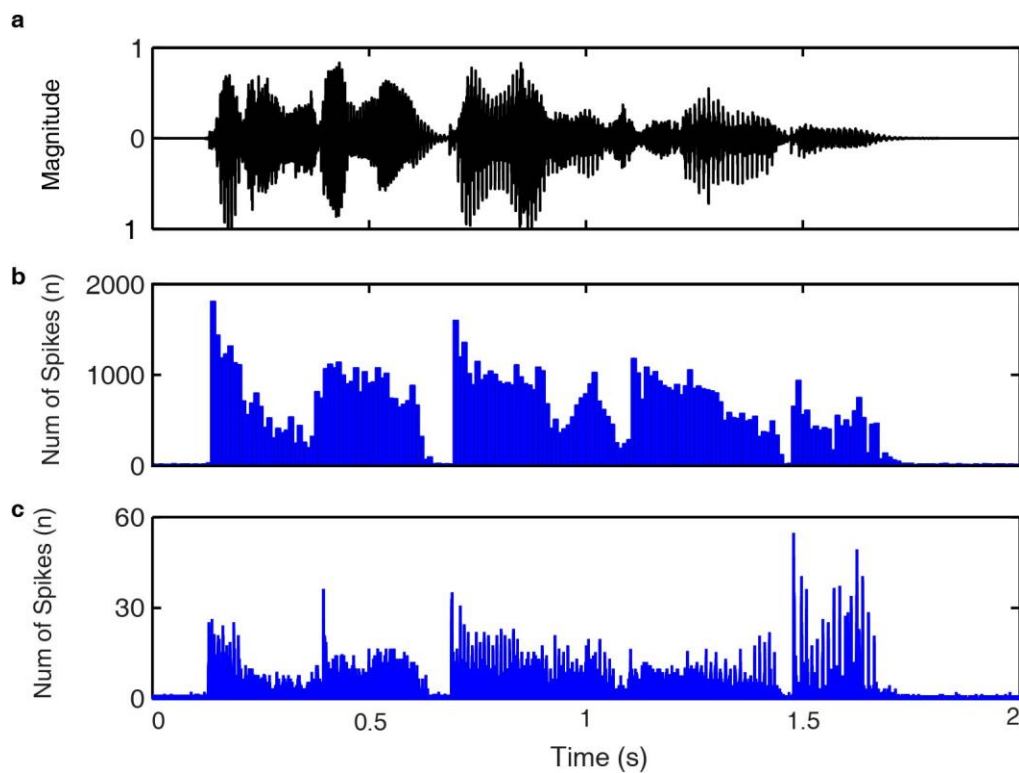
Figure 5- 2. The time waveform (a) was shown for the speech sentence "the silly boy's hiding". Two PSTHs (b and c) from an individual ICC neuron showed the neural response to the given speech sentence. The bin size of PSTH was 10 ms (b) and 0.25 ms (c) respectively. This histogram was acquired with 100 repetitions.

A direct comparison of the two PSTHs showed that PSTH with a small bin size shows more fine structure information than the PSTH with a large bin size (Figure 5-3).

Figure 5- 3. Speech signal and neural PSTH.
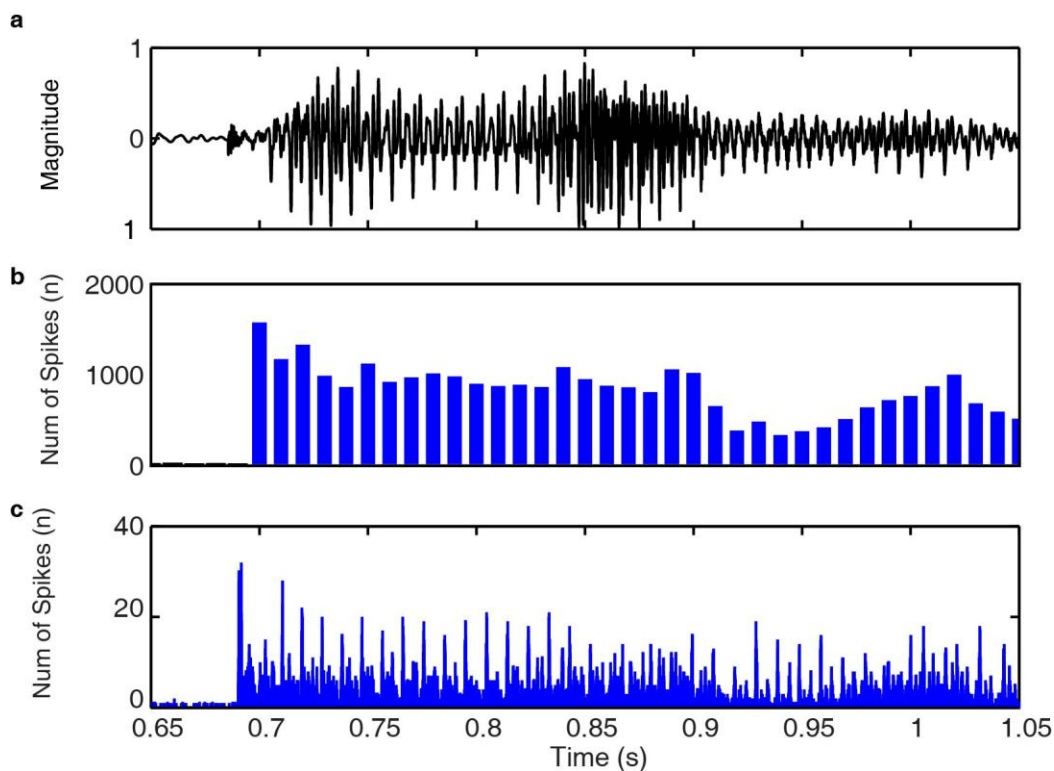
Figure 5- 3. The time waveform (a) was shown for a 0.4-s section of speech sentence "the silly boy's hiding" from 0.65 to 1.05 s. Two PSTHs (b and c) from a single ICC neuron show the neural response to the given speech sentence. The bin size of PSTH was 10 ms (b) and 0.25 ms (c) respectively. This histogram was acquired with 100 repetitions. The sampling rate of the recording system used in the experiment was 40,000 Hz. The PSTHs used for spectral analysis and which results are shown in the manuscript were not binned, taking advantage of the 40,000 Hz sampling rate. Thus, the spectral analysis of the PSTH can reveal neural synchrony of acoustic signal up to 20,000 Hz. Of note, there were 100 repetitions for each recording.

5.2.2 Temporal features of natural speech: its neural representation and percental function

Chapter 4 of this thesis focused on the neural process of temporal cues in ICC units. We have examined the systematically characterized the firing patterns of individual neurons from the inferior colliculus in response to natural English sentences. The data showed that 74% of ICC

neurons showed a phase locking to the F0. Additionally, the low frequency speech envelope was coded by the average rate of neural discharges in up to 67% of the neurons of the central nucleus of the inferior colliculus (ICC). The ICC neurons can follow the time-varying F0 in natural English speech. Our result suggested that the predominant temporal coding in the periphery degrades and transits to the rate coding in the cortex [130], which has likely taken the place in IC.

Chapter 3 examined the perceptual function of speech temporal cues. Our data showed that that speech intelligibility in humans relied on the periodic cues of speech TFS in both quiet and noisy listening conditions. Furthermore, recordings of neural activity from the guinea pig inferior colliculus have shown that individual ICC units exhibit phase locking patterns to the periodic cues of speech TFS that disappear when reconstructed sounds do not show periodic patterns anymore. Thus, the periodic cues of TFS are essential for speech intelligibility and are encoded in auditory neurons by phase locking.

Future directions lie in understanding the perceptual role and neural coding of periodic patterns in subjects with different levels and types of hearing loss, as well as how to encode temporal cues in electrical and optical stimulations for better performances. Studies have suggested that listeners with sensorineural hearing loss may have an increase in the perceptual salience of envelope structure that could adversely affect speech perception in fluctuating background noise through loudness recruitment[214-216]. People with moderate flat hearing loss showed a decreased ability to use TFS cues[195]. It has also been suggested that the temporal precision of E coding could be enhanced with hearing loss at an equal stimulus sensation level in chinchillas[217].

In addition, pronounced distortions in the tonotopic coding of temporal cues in auditory nerve fibers have been found in chinchilla after noise-induced hearing loss [218].

5.3 Summary

This thesis focuses on the development of a cochlear implant (CI) that uses photons to stimulate surviving auditory neurons in severe-to-profound deaf individuals. The benefit of optical stimulation is its spatial selectivity with the potential to create significantly more independent channels to encode acoustic information that likely enhances the CI users' performance in challenging listening environments. We have developed a light delivery system using small light sources that evoked auditory response. Additionally, we have demonstrated that the temporal periodicity is essential for speech perception in normal hearing listeners in both quiet and noisy environment. We have also discussed the challenges and pointed discussed research directions towards the optical cochlear implants.

References

[1]   Webb K, Connor S, Wilson K, Cooper S and Jiang D 2015 Tough choices: The challenges of cochlear implantation when there is 'something to lose' *Cochlear Implants International* **16 Suppl 1** S50-2

[2]   Bierer J A and Middlebrooks J C 2002 Auditory cortical images of cochlear-implant stimuli: dependence on electrode configuration *Journal of neurophysiology* **87** 478-92

[3]   Everest F A 2001 Master handbook of acoustics. ASA)

[4]   Borg E 1973 On the neuronal organization of the acoustic middle ear reflex. A physiological and anatomical study *Brain research* **49** 101-23

[5]   Guinan J J, Jr., Backus B C, Lilaonitkul W and Aharonson V 2003 Medial olivocochlear efferent reflex in humans: otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs *Journal of the Association for Research in Otolaryngology : JARO* **4** 521-40

[6]   Mukerji S, Windsor A M and Lee D J 2010 Auditory brainstem circuits that mediate the middle ear muscle reflex *Trends Amplif* **14** 170-91

[7]   Pickles J O 2008 An introduction to the physiology of hearing

[8]   Clark J G 1981 Uses and abuses of hearing loss classification *Asha* **23** 493-500

[9]   Picton T W, Durieux-Smith A and Moran L M 1994 Recording auditory brainstem responses from infants *International journal of pediatric otorhinolaryngology* **28** 93-110

[10] Wilson B S and Dorman M F 2008 Cochlear implants: current designs and future possibilities *Journal of rehabilitation research and development* **45** 695-730

[11] Battmer R D, Haake P, Zilberman Y and Lenarz T 1999 Simultaneous analog stimulation (SAS)–continuous interleaved sampler (CIS) pilot comparison study in Europe *Annals of Otology, Rhinology & Laryngology* **108** 69-73

[12] Osberger M J and Fisher L 2000 New directions in speech processing: patient performance with simultaneous analog stimulation *Annals of Otology, Rhinology & Laryngology* **109** 70-3

[13] Loizou P C, Stickney G, Mishra L and Assmann P 2003 Comparison of speech processing strategies used in the Clarion implant processor *Ear and hearing* **24** 12-9

[14] Wilson B S, Finley C C, Lawson D T, Wolford R D, Eddington D K and Rabinowitz W M 1991 Better speech recognition with cochlear implants *Nature* **352** 236-8

[15] McDermott H J, Mckay C M and Vandali A E 1992 A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant *The Journal of the Acoustical Society of America* **91** 3367-71

[16] Lawson D, Wilson B, Zerbi M and Finley C 1996 Speech processors for auditory prostheses (Third quarterly progress report, NIH project N01-DC-5–2103) *National Institutes of Health, Bethesda, MD: Neural prostheses program*

[17] Ziese M, Stützel A, von Specht H, Begall K, Freigang B, Sroka S and Nopp P 2000 Speech understanding with the CIS and the n-of-m strategy in the MED-EL COMBI 40+ system *ORL* **62** 321-9

[18]   Arndt P 1999 Within subject comparison of advanced coding strategies in the Nucleus 24 cochlear implant. In: *1999 Conference on Implantable Auditory Prostheses, Asilomar, CA,*

[19]   Vandali A E, Whitford L A, Plant K L and Clark G M 2000 Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system *Ear and hearing* **21** 608-24

[20]   Kiefer J, Hohl S, Stürzebecher E, Pfennigdorff T and Gstöettner W 2001 Comparison of Speech Recognition with Different Speech Coding Strategies (SPEAK, CIS, and ACE) and Their Relationship to Telemetric Measures of Compound Action Potentials in the Nucleus CI 24M Cochlear Implant System: Comparación del reconocimiento del lenguaje utilizando diferentes estrategias (SPEAK, CIS y ACE) y su relación con mediciones telemétricas de potenciales de acción compuestos, con el sistema de implante coclear nucleus CI24M *Audiology : official organ of the International Society of Audiology* **40** 32-42

[21]   Skinner M W, Clark G M, Whitford L A, Seligman P M, Staller S J, Shipp D B, Shallop J K, Everingham C, Menapace C M and Arndt P L 1994 Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system *Scientific publications, vol. 8, 1994-1995, no. 685*

[22]   Seligman P and McDermott H 1995 Architecture of the Spectra 22 speech processor *Scientific publications, vol. 8, 1994-1995, no. 761*

[23]   Helms J, Müller J, Schön F, Winkler F, Moser L, Shehata-Dieler W, Kastenbauer E, Baumann U, Rasp G and Schorn K 2001 Comparison of the TEMPO+ ear-level speech

processor and the CIS PRO+ body-worn processor in adult MED-EL cochlear implant users *ORL* **63** 31-40

[24]   Eddington D, Dobelle W H, Mladejovsky M, Brackmann D and Parkin J 1978 Auditory prostheses research with multiple channel intracochlear stimulation in man *Annals of Otology, Rhinology & Laryngology* **87** 5-39

[25]   Zeng F-G and Shannon R V 1992 Loudness balance between electric and acoustic stimulation *Hearing research* **60** 231-5

[26]   Dorman M F 1993 Speech perception by adults *Cochlear implants audiological foundations. San Diego, CA, USA: Singular Publishing Group*

[27]   Favre E and Pelizzone M 1993 Channel interactions in patients using the Ineraid multichannel cochlear implant *Hearing research* **66** 150-6

[28]   Shannon R V 1983 Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics *Hearing research* **11** 157-89

[29]   Tong Y, Blamey P, Dowell R and Clark G M 1983 Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant *The Journal of the Acoustical Society of America* **74** 73-80

[30]   Townshend B, Cotter N, Van Compernolle D and White R 1987 Pitch perception by cochlear implant subjects *The Journal of the Acoustical Society of America* **82** 106-15

[31]   Zeng F-G 2002 Temporal pitch in electric hearing *Hearing research* **174** 101-6

[32]   Battmer R, Kuzma J, Frohne C, Goldring J and Lenarz T 1998 Better modiolus hugging electrode placement: electrophysiological and clinical result of new Clarion electrode positioner. In: *Proceedings of the 4th European symposium on pediatric cochlear implantation,*

[33]   Kuzma J A 1996 Cochlear electrode implant assemblies with positioning system therefor. Google Patents)

[34]   Spelman F A, Clopton B M, Voie A, Jolly C N, Huynh K, Boogaard J and Swanson J W 1998 Cochlear implant with shape memory material and method for implanting the same. Google Patents)

[35]   Snyder R L, Bierer J A and Middlebrooks J C 2004 Topographic spread of inferior colliculus activation in response to acoustic and intracochlear electric stimulation *JARO* **5** 305-22

[36]   Mens L H and Berenstein C K 2005 Speech perception with mono- and quadrupolar electrode configurations: a crossover study *Otol Neurotol* **26** 957-64

[37]   Srinivasan A G, Landsberger D M and Shannon R V 2010 Current focusing sharpens local peaks of excitation in cochlear implant stimulation *Hearing research* **270** 89-100

[38]   Wu C C and Luo X 2016 Excitation Patterns of Standard and Steered Partial Tripolar Stimuli in Cochlear Implants **17** 145-58

[39]   Koch D B, Downing M, Osberger M J and Litvak I M 2006 Using current steering to increase spatial resolution in CII and HiRes 90K users. *Ear and hearing* (in press)

[40]   Berenstein C K, Mens L H, Mulder J J and Vanpoucke F J 2008 Current steering and current focusing in cochlear implants: comparison of monopolar, tripolar, and virtual channel electrode configurations **29** 250-60

[41]   Choi C T and Hsu C H 2009 Conditions for generating virtual channels in cochlear prosthesis systems **37** 614-24

[42]   Landsberger D M and Srinivasan A G 2009 Virtual channel discrimination is improved by current focusing in cochlear implant recipients **254** 34-41

[43]   van den Honert C and Kelsall D C 2007 Focused intracochlear electric stimulation with phased array channels *The Journal of the Acoustical Society of America* **121** 3703-16

[44]   Izzo A D, Richter C-P, Jansen E D and Walsh J T 2006 Laser stimulation of the auditory nerve *Laser Surg Med* **38** 745-53

[45]   Hernandez V H, Gehrt A, Reuter K, Jing Z, Jeschke M, Mendoza Schulz A, Hoch G, Bartels M, Vogt G, Garnham C W, Yawo H, Fukazawa Y, Augustine G J, Bamberg E, Kugler S, Salditt T, de Hoz L, Strenzke N and Moser T 2014 Optogenetic stimulation of the auditory pathway *J Clin Invest* **124** 1114-29

[46]   Richter C P and Tan X 2014 Photons and neurons *Hearing research* **311** 72-88

[47]   Wells J, Kao C, Jansen E D, Konrad P and Mahadevan-Jansen A 2005 Application of infrared light for in vivo neural stimulation **10** 064003

[48]   Wells J, Kao C, Mariappan K, Albea J, Jansen E D, Konrad P and Mahadevan-Jansen A 2005 Optical stimulation of neural tissue in vivo *Optics letters* **30** 504-6

[49]   Moreno L E, Rajguru S M, Matic A I, Yerram N, Robinson A M, Hwang M, Stock S and Richter C P 2011 Infrared neural stimulation: beam path in the guinea pig cochlea *Hearing research* **282** 289-302

[50]   Banakis R M, Matic A I, Rajguru S M and Richter C-P 2011 Optical stimulation of the auditory nerve: effects of pulse shape. In: *Photonic Therapeutics and Diagnostics VII*: International Society for Optics and Photonics) p 788358

[51]   Goyal V, Rajguru S, Matic A I, Stock S R and Richter C P 2012 Acute damage threshold for infrared neural stimulation of the cochlea: functional and histological evaluation *Anatomical record* **295** 1987-99

[52]   Izzo A D, Richter C P, Jansen E D and Walsh J T, Jr. 2006 Laser stimulation of the auditory nerve *Lasers Surg Med* **38** 745-53

[53]   Izzo A D, Suh E, Pathria J, Walsh J T, Whitlon D S and Richter C-P 2007 Selectivity of neural stimulation in the auditory system: a comparison of optic and electric stimuli *Journal of biomedical optics* **12** 021008

[54]   Albert E S, Bec J M, Desmadryl G, Chekroud K, Travo C, Gaboyard S, Bardin F, Marc I, Dumas M, Lenaers G, Hamel C, Muller A and Chabbert C 2012 TRPV4 channels mediate the infrared laser-evoked response in sensory neurons *Journal of neurophysiology* **107** 3227-34

[55]   Yao J, Liu B and Qin F 2009 Rapid temperature jump by infrared diode laser irradiation for patch-clamp studies *Biophysical journal* **96** 3611-9

[56] Rhee A Y, Li G, Wells J and Kao Y P Y 2008 Photostimulation of sensory neurons of the rat vagus nerve. *SPIE* **6854** 68540E1

[57] Suh E, Matic A I, Otting M, Walsh Jr. J T and Richter C-P 2009 Optical stimulation in mice which lack the TRPV1 channel. *Proc. of SPIE* **7180** 71800S 1-5

[58] Güler A D, Lee H, Iida T, Shimizu I, Tominaga M and Caterina M 2002 Heat-evoked activation of the ion channel, TRPV4 *The Journal of neuroscience : the official journal of the Society for Neuroscience* **22** 6408-14

[59] Sladek C D and Johnson A K 2013 Integration of thermal and osmotic regulation of water homeostasis: the role of TRPV channels *Am J Physiol Regul Integr Comp Physiol* **305** R669-78

[60] Santoni G, Farfariello V and Amantini C 2011 TRPV channels in tumor growth and progression *Adv Exp Med Biol* **704** 947-67

[61] Baylie R L and Brayden J E 2011 TRPV channels and vascular function *Acta Physiol (Oxf)* **203** 99-116

[62] Kauer J A and Gibson H E 2009 Hot flash: TRPV channels in the brain *Trends in neurosciences* **32** 215-24

[63] Sharif-Naeini R, Ciura S, Zhang Z and Bourque C W 2008 Contribution of TRPV channels to osmosensory transduction, thirst, and vasopressin release *Kidney Int* **73** 811-5

[64] Jia Y and Lee L Y 2007 Role of TRPV receptors in respiratory diseases *Biochim Biophys Acta* **1772** 915-27

[65]   Lee H and Caterina M J 2005 TRPV channels as thermosensory receptors in epithelial cells *Pflugers Arch* **451** 160-7

[66]   O'Neil R G and Heller S 2005 The mechanosensitive nature of TRPV channels *Pflugers Arch* **451** 193-203

[67]   Gunthorpe M J, Benham C D, Randall A and Davis J B 2002 The diversity in the vanilloid (TRPV) receptor family of ion channels *Trends Pharmacol Sci* **23** 183-91

[68]   Rajguru S M, Rabbitt R R, Matic A I, Highstein S M and Richter C P 2010 Inhibitory and Excitatory Vestibular Afferent Responses Induced By Infrared Light Stimulation of Hair Cells. In: *33rd Midwinter Meeting,* (Anaheim, CA: Association for Research in Otolaryngology)

[69]   Rajguru S M, Richter C P, Matic A I, Holstein G R, Highstein S M, Dittami G M and Rabbitt R D 2011 Infrared photostimulation of the crista ampullaris *The Journal of physiology* **589** 1283-94

[70]   Lumbreras V, Finale M, Bas E, Gupta C and Rajguru S 2013 Pulsed infrared-evoked intracellular calcium transients in cultured neonatal spiral ganglion neurons. *Abstr. Assoc. Res. Otolaryngol.* **36** 341

[71]   Lumbreras V, Bas E, Gupta C and Rajguru S M 2014 Pulsed infrared radiation excites cultured neonatal spiral and vestibular ganglion neurons by modulating mitochondrial calcium cycling *Journal of neurophysiology* **112** 1246-55

[72]   Teudt I U, Maier H, Richter C P and Kral A 2011 Acoustic events and "optophonic" cochlear responses induced by pulsed near-infrared laser *IEEE transactions on bio-medical engineering* **58** 1648-55

[73]   Schultz M, Baumhoff P, Maier H, Teudt I U, Kruger A, Lenarz T and Kral A 2012 Nanosecond laser pulse stimulation of the inner ear-a wavelength study *Biomed Opt Express* **3** 3332-45

[74]   Schultz M, Baumhoff P, Teudt I U, Maier H, Kruger A, Lenarz T and Kral A 2012 Pulsed wavelength-dependent laser stimulation of the inner ear *Biomed Tech (Berl)* **57 Suppl 1**

[75]   Thompson A C, Fallon J B, Wise A K, Wade S A, Shepherd R K and Stoddart P R 2015 Infrared neural stimulation fails to evoke neural activity in the deaf guinea pig cochlea *Hearing research* **324** 46-53

[76]   Richter C-P, Bayon R, Izzo A D, Otting M, Suh E, Goyal S, Hotaling J and Walsh Jr. J T 2008 Optical stimulation of auditory neurons: effects of acute and chronic deafening *Hearing research* **242** 42-51

[77]   Tan X, Rajguru S, Young H, Xia N, Stock S R, Xiao X and Richter C P 2015 Radiant energy required for infrared neural stimulation *Scientific reports* **5** 13273

[78]   Young H K, Tan X, Xia N and Richter C P 2015 Target structures for cochlear infrared neural stimulation *Neurophotonics* **2** 025002

[79]   Tan X, Jahan I, Xu Y, Stock S, Claire K, Soriano C, Xiao X, García-Añoveros J, Fritzsch B and Richter C-P 2018 Auditory Neural Activity in Congenitally Deaf Mice Induced by Infrared Neural Stimulation. *Scientific reports* **8** 388

[80]   Obholzer N, Wolfson S, Trapani J G, Mo W, Nechiporuk A, Busch-Nentwich E, Seiler C, Sidi S, Sollner C, Duncan R N, Boehland A and Nicolson T 2008 Vesicular glutamate transporter 3 is required for synaptic transmission in zebrafish hair cells *The Journal of neuroscience : the official journal of the Society for Neuroscience* **28** 2110-8

[81]   Seal R P, Akil O, Yi E, Weber C M, Grant L, Yoo J, Clause A, Kandler K, Noebels J L, Glowatzki E, Lustig L R and Edwards R H 2008 Sensorineural deafness and seizures in mice lacking vesicular glutamate transporter 3 *Neuron* **57** 263-75

[82]   Ahnert-Hilger G and Jahn R 2008 Into great silence without VGLUT3 *Neuron* **57** 173-4

[83]   Jahan I, Pan N, Kersigo J and Fritzsch B 2015 Neurog1 can partially substitute for Atoh1 function in hair cell differentiation and maintenance during organ of Corti development *Development* **142** 2810-21

[84]   Jahan I, Pan N, Kersigo J, Calisto L E, Morris K A, Kopecky B, Duncan J S, Beisel K W and Fritzsch B 2012 Expression of Neurog1 instead of Atoh1 can partially rescue organ of Corti cell survival *PloS one* **7** e30853

[85]   Balster S, Wenzel G I, Warnecke A, Steffens M, Rettenmaier A, Zhang K, Lenarz T and Reuter G 2014 Optical cochlear implant: evaluation of insertion forces of optical fibres in a cochlear model and of traumata in human temporal bones **59** 19-28

[86]   Galambos R and Davis H 1943 The response of single auditory-nerve fibers to acoustic stimulation *Journal of neurophysiology* **6** 39-57

[87]   Rao A and Carney L H 2014 Speech enhancement for listeners with hearing loss based on a model for vowel coding in the auditory midbrain *IEEE transactions on bio-medical engineering* **61** 2081-91

[88]   Carney L H, Li T and McDonough J M 2015 Speech Coding in the Brain: Representation of Vowel Formants by Midbrain Neurons Tuned to Sound Fluctuations(1,2,3) *eNeuro* **2**

[89]   Carney L H, Kim D O and Kuwada S 2016 Speech Coding in the Midbrain: Effects of Sensorineural Hearing Loss *Adv Exp Med Biol* **894** 427-35

[90]   Richter C-P, LaFaire P and Tan X 2016 System and method for animal-human neural interface. Google Patents)

[91]   Rosen S 1992 Temporal information in speech: acoustic, auditory and linguistic aspects *Philos Trans R Soc Lond B Biol Sci* **336** 367-73

[92]   Langner G 1992 Periodicity coding in the auditory system *Hearing research* **60** 115-42

[93]   Green T, Faulkner A, Rosen S and Macherey O 2005 Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification *The Journal of the Acoustical Society of America* **118** 375

[94]   Steinmetzger K and Rosen S 2015 The role of periodicity in perceiving speech in quiet and in background noise *The Journal of the Acoustical Society of America* **138** 3586-99

[95]   Steinmetzger K and Rosen S 2017 Effects of acoustic periodicity and intelligibility on the neural oscillations in response to speech *Neuropsychologia* **95** 173-81

[96]   Steinmetzger K and Rosen S 2017 Effects of acoustic periodicity, intelligibility, and pre-stimulus alpha power on the event-related potentials in response to speech *Brain Lang* **164** 1-8

[97]   Young E D and Sachs M B 1979 Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers *The Journal of the Acoustical Society of America* **66** 1381-403

[98]   Johnson D H 1980 The Relationship between Spike Rate and Synchrony in Responses of Auditory-Nerve Fibers to Single Tones *Journal of the Acoustical Society of America* **68** 1115-22

[99]   Palmer A R and Russell I J 1986 Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells *Hearing research* **24** 1-15

[100]   Liu L F, Palmer A R and Wallace M N 2006 Phase-locked responses to pure tones in the inferior colliculus *Journal of neurophysiology* **95** 1926-35

[101]   Delgutte B 1984 Speech coding in the auditory nerve: I. Vowel-like sounds *The Journal of the Acoustical Society of America* **75** 866

[102]   Delgutte B 1984 Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds *The Journal of the Acoustical Society of America* **75** 879

[103]   Delgutte B 1984 Speech coding in the auditory nerve: III. Voiceless fricative consonants *The Journal of the Acoustical Society of America* **75** 887

[104]   Delgutte B 1984 Speech coding in the auditory nerve: V. Vowels in background noise *The Journal of the Acoustical Society of America* **75** 908

[105]   Delgutte B 1984 Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics *The Journal of the Acoustical Society of America* **75** 897

[106]   Pfeiffer R R, Molnar C E and Cox J R 1974 The Representation of Tones and Combination Tones in Spike Discharge Patterns of Single Cochlear Nerve Fibers **8** 323-31

[107]   Goldstein R, Rodman L B and Karlovich R S 1972 Effects of Stimulus Rate and Number on the Early Components of the Averaged Electroencephalic Response *Journal of Speech Language and Hearing Research* **15** 559

[108]   Jackson H M and Moore B C 2013 The dominant region for the pitch of complex tones with low fundamental frequencies *The Journal of the Acoustical Society of America* **134** 1193-204

[109]   Remez R E, Rubin P E, Pisoni D B and Carrell T D 1981 Speech perception without traditional speech cues *Science* **212** 947-9

[110]   Shannon R V, Zeng F G, Kamath V, Wygonski J and Ekelid M 1995 Speech recognition with primarily temporal cues *Science* **270** 303-4

[111]   Kazama M, Gotoh S, Tohyama M and Houtgast T 2010 On the significance of phase in the short term Fourier spectrum for speech intelligibility *The Journal of the Acoustical Society of America* **127** 1432-9

[112]   Dubbelboer F and Houtgast T 2007 A detailed study on the effects of noise on speech intelligibility *The Journal of the Acoustical Society of America* **122** 2865-71

[113]   Joris P, Schreiner C and Rees A 2004 Neural processing of amplitude-modulated sounds *Physiol Rev* **84** 541-77

[114]   Xu Y, Chen M, LaFaire P, Tan X and Richter C-P 2017 Distorting temporal fine structure by phase shifting and its effects on speech intelligibility and neural phase locking *Scientific reports* **7**

[115]   Plomp R 1967 Pitch of complex tones *The Journal of the Acoustical Society of America* **41** 1526-33

[116]   Rose J E, Brugge J F, Anderson D J and Hind J E 1967 Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey *Journal of neurophysiology* **30** 769-93

[117]   Hind J E, Anderson D J, Brugge J F and Rose J E 1967 Coding of information pertaining to paired low-frequency tones in single auditory nerve fibers of the squirrel monkey *Journal of neurophysiology* **30** 794-816

[118]   Johnson D H 1974 The response of single auditory-nerve fibers in the cat to single tones: synchrony and average discharge rate. Massachusetts Institute of Technology)

[119]   Johnson D and Kiang N 1976 Analysis of discharges recorded simultaneously from pairs of auditory nerve fibers *Biophysical journal* **16** 719-34

[120] Joris P X, Carney L H, Smith P H and Yin T 1994 Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency *Journal of neurophysiology* **71** 1022-36

[121] Köppl C 1997 Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, Tyto alba *Journal of Neuroscience* **17** 3312-21

[122] Winter I M and Palmer A R 1990 Responses of single units in the anteroventral cochlear nucleus of the guinea pig *Hearing research* **44** 161-78

[123] Goldberg J M and Brownell. W E 1973 Discharge charateristics of neurons in anteroventral and dorsal cochlear nuclei of cat *Brain research* **64** 35-54

[124] Langner G and Schreiner C E 1988 Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms *Journal of neurophysiology* **60** 1799-822

[125] Schreiner C E and Langner G 1988 Periodicity coding in the inferior colliculus of the cat. II. Topographical organization *Journal of neurophysiology* **60** 1823-40

[126] Bendor D and Wang X 2005 The neuronal representation of pitch in primate auditory cortex *Nature* **436** 1161-5

[127] Winter I M, Palmer A R, Wiegrebe L and Patterson R D 2003 Temporal coding of the pitch of complex sounds by presumed multipolar cells in the ventral cochlear nucleus *Speech Communication* **41** 135-49

[128] Cariani P A and Delgutte B 1996 Neural correlates of the pitch of complex tones. I. Pitch and pitch salience *Journal of neurophysiology* **76** 1698-716

[129] Cedolin L and Delgutte B 2005 Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve *Journal of neurophysiology* **94** 347-62

[130] Oxenham A J 2018 How We Hear: The Perception and Neural Coding of Sound *Annual review of psychology* **69**

[131] Schreiner C E and Urbas J V 1986 Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF) *Hearing research* **21** 227-41

[132] Schreiner C E and Urbas J V 1988 Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields *Hearing research* **32** 49-63

[133] Su Y and Delgutte B 2018 Pitch of Harmonic Complex Tones: Rate Coding of Envelope Repetition Rate in the Auditory Midbrain *Acta Acustica united with Acustica* **104** 860-4

[134] De Ribaupierre F, Rouiller E, Toros A and De Ribaupierre Y 1980 Transmission delay of phase-locked cells in the medial geniculate body *Hearing research* **3** 65-77

[135] Aitkin L and Webster W 1972 Medial geniculate body of the cat: organization and responses to tonal stimuli of neurons in ventral division *Journal of neurophysiology* **35** 365-80

[136] Wallace M N, Rutkowski R G, Shackleton T M and Palmer A R 2000 Phase-locked responses to pure tones in guinea pig auditory cortex *Neuroreport* **11** 3989-93

[137] Wallace M N, Shackleton T M and Palmer A R 2002 Phase-locked responses to pure tones in the primary auditory cortex *Hearing research* **172** 160-71

[138] Sullivan W and Konishi M 1984 Segregation of stimulus phase and intensity coding in the cochlear nucleus of the barn owl *Journal of Neuroscience* **4** 1787-99

[139] Palmer A, Winter I and Darwin C 1986 The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons *The Journal of the Acoustical Society of America* **79** 100-13

[140] Weiss T and Rose C 1988 A comparison of synchronization filters in different auditory receptor organs *Hearing research* **33** 175-9

[141] Javel E 1980 Coding of AM tones in the chinchilla auditory nerve: implications for the pitch of complex tones *The Journal of the Acoustical Society of America* **68** 133-46

[142] Joris P X and Yin T C 1992 Responses to amplitude-modulated tones in the auditory nerve of the cat *The Journal of the Acoustical Society of America* **91** 215-32

[143] Rose G J and Capranica R R 1985 Sensitivity to amplitude modulated sounds in the anuran auditory nervous system *Journal of neurophysiology* **53** 446-65

[144] Palmer A, Rees A and Caird D 1990 Interaural delay sensitivity to tones and broad band signals in the guinea-pig inferior colliculus *Hearing research* **50** 71-86

[145] Joris P X, Van De Sande B and van der Heijden M 2005 Temporal damping in response to broadband noise. I. Inferior colliculus *Journal of neurophysiology* **93** 1857-70

[146] Deng L and Geisler C D 1987 Responses of auditory-nerve fibers to nasal consonant–vowel syllables *The Journal of the Acoustical Society of America* **82** 1977-88

[147]   Miller M I and Sachs M B 1984 Representation of voice pitch in discharge patterns of auditory-nerve fibers *Hearing research* **14** 257-79

[148]   Peng F, Innes-Brown H, McKay C M, Fallon J B, Zhou Y, Wang X, Hu N and Hou W 2018 Temporal Coding of Voice Pitch Contours in Mandarin Tones *Frontiers in neural circuits* **12** 55

[149]   Nogueira W, Litvak L M, Landsberger D M and Büchner A 2017 Loudness and pitch perception using Dynamically Compensated Virtual Channels **344** 223-34

[150]   Moser T, Hernandez V H, Hoch G, Reuter K, Jing Z, Bartels M, Vogt G, Garnham C W, Augustine G J, Kügler S, Salditt T and Strenzke N 2013 Optogenetic Stimulation of the Auditory Nerve. *Abstr. Assoc. Res. Otolaryngol.* **36** 268

[151]   Zheng J F, Dai C F, Steyger P S, Kim Y, Vass Z, Ren T Y and Nuttall A L 2003 Vanilloid receptors in hearing: Altered cochlear sensitivity by vanilloids and expression of TRPV1 in the organ of Corti *Journal of neurophysiology* **90** 444-55

[152]   Hernandez V H, Gehrt A, Reuter K, Jing Z, Jeschke M, Schulz A M, Hoch G, Bartels M, Vogt G, Garnham C W, Yawo H, Fukazawa Y, Augustine G J, Bamberg E, Kuugler S, Salditt T, de Hoz L, Strenzke N and Moser T 2014 Optogenetic stimulation of the auditory pathway *Journal of Clinical Investigation* **124** 1114-29

[153]   Wells J, Kao C, Konrad P, Milner T, Kim J, Mahadevan-Jansen A and Jansen E D 2007 Biophysical mechanisms of transient optical stimulation of peripheral nerve *Biophysical journal* **93** 2567-80

[154]   Liu Q, Frerck M J, Holman H A, Jorgensen E M and Rabbitt R D 2014 Exciting cell membranes with a blustering heat shock *Biophysical journal* **106** 1570-7

[155]   Rabbitt R D, Brichta A M, Tabatabaee H, Boutros P J, Ahn J, Della Santina C C, Poppi L A and Lim R 2016 Heat pulse excitability of vestibular hair cells and afferent neurons *Journal of neurophysiology* **116** 825-43

[156]   Xia N, Tan X, Xu Y, Hou W, Mao T and Richter C P 2016 Pressure in the cochlea during infrared irradiation *IEEE transactions on bio-medical engineering*

[157]   Thompson A C, Wade S A, Brown W G and Stoddart P R 2012 Modeling of light absorption in tissue during infrared neural stimulation *Journal of Biomedical Optics* **17** 075002

[158]   Thompson A C, Wade S A, Cadusch P J, Brown W G and Stoddart P R 2013 Modeling of the temporal effects of heating during infrared neural stimulation *Journal of Biomedical Optics* **18** 035004

[159]   Thompson A C, Wade S A, Pawsey N C and Stoddart P R 2013 Infrared Neural Stimulation: Influence of Stimulation Site Spacing and Repetition Rates on Heating *IEEE Transactions on Biomedical Engineering* **60** 3534-41

[160]   Keller D M, Wells J D, Dummer M and Hibbs-Brenner M 2012 Novel hardware development for infrared neural stimulation *SPIE* **8207G-153**

[161]   Niemz M H 2004 *Laser Tissue interactions: fundamentals and application.* (New York: Springer)

[162]   Welch A J and van Gemert M J C 2012 *Optical-Thermal Response of Laser-Irradiated Tissue.* (New York: Plenum Press)

[163]   Fork R L 1971 Laser stimulation of nerve cells in Aplysia *Science* **171** 907-8

[164]   Shapiro M G, Homma K, Villarreal S, Richter C P and Bezanilla F 2012 Infrared light excites cells by changing their electrical capacitance *Nature communications* **3** 736

[165]   Liu Q, Jorgensen E, Holman H, Frerck M and Rabbitt R D 2013 Miniature post synaptic currents are entrained by infrared pulses. *Abstr. Assoc. Res. Otolaryngol.* 464

[166]   Okunade O and Santos-Sacchi J 2013 IR laser-induced perturbations of the voltage-dependent solute carrier protein SLC26a5 *Biophysical journal* **105** 1822-8

[167]   Rabbitt R D, Lim R, Tabatabaee H, Poppi L, Ferek M and Brichta A 2016 Excitation and inhibition of semicircular canal type II hair cells by pulsed infrared light. *Abstr. Assoc. Res. Otolaryngol.* **39, PS64**

[168]   Plaksin M, Kimmel E and Shoham S 2017 Thermal transients excite neurons through universal intramembrane mechano-electrical effects. *bioRxiv 111724*

[169]   Beier H T, Tolstykh G P, Musick J D, Thomas R J and Ibey B L 2014 Plasma membrane nanoporation as a possible mechanism behind infrared excitation of cells *Journal of neural engineering* **11** 066006

[170]   Moreau D, Lefort C, Pas J, Bardet S M, Leveque P and O'Connor R P 2018 Infrared neural stimulation induces intracellular Ca2+ release mediated by phospholipase C **11**

[171] Stakhovskaya O, Sridhar D, Bonham B H and Leake P A 2007 Frequency map for the human cochlear spiral ganglion: implications for cochlear implants **8** 220-33

[172] Landsberger D M, Svrakic M, Roland J T, Jr. and Svirsky M 2015 The Relationship Between Insertion Angles, Default Frequency Allocations, and Spiral Ganglion Place Pitch in Cochlear Implants *Ear & Hearing* **36** e207-13

[173] Ejserholm F, Stegmayr J, Bauer P, Johansson F, Wallman L, Bengtsson M and Oredsson S 2015 Biocompatibility of a polymer based on Off-Stoichiometry Thiol-Enes + Epoxy (OSTE+) for neural implants *Biomater Res* **19** 19

[174] Mattioli-Belmonte M, Giavaresi G, Biagini G, Virgili L, Giacomini M, Fini M, Giantomassi F, Natali D, Torricelli P and Giardino R 2003 Tailoring biomaterial compatibility: in vivo tissue response versus in vitro cell behavior *Int J Artif Organs* **26** 1077-85

[175] Sun Y, Lacour S P, Brooks R A, Rushton N, Fawcett J and Cameron R E 2009 Assessment of the biocompatibility of photosensitive polyimide for implantable medical device use *J Biomed Mater Res A* **90** 648-55

[176] Starr P, Agrawal C M and Bailey S 2016 Biocompatibility of common polyimides with human endothelial cells for a cardiovascular microsensor *J Biomed Mater Res A* **104** 406-12

[177] Bae S H, Che J H, Seo J M, Jeong J, Kim E T, Lee S W, Koo K I, Suaning G J, Lovell N H, Cho D I, Kim S J and Chung H 2012 In vitro biocompatibility of various polymer-based microelectrode arrays for retinal prosthesis *Invest Ophthalmol Vis Sci* **53** 2653-7

[178] Avci E, Nauwelaers T, Lenarz T, Hamacher V and Kral A 2014 Variations in microanatomy of the human cochlea **522** 3245-61

[179] Erixon E, Högstorp H, Wadin K and Rask-Andersen H 2009 Variational anatomy of the human cochlea: implications for cochlear implantation. *Otology & Neurotology* **30** 14-22

[180] Salt A 2017 http://oto2wustl.edu/cochlea/mrhmvol.htm.

[181] Hatsushika S, Shepherd R K, Tong Y C, Clark G M and Funasaka S 1990 Dimensions of the scala tympani in the human and cat with reference to cochlear implants **99** 871-6

[182] Matic A I, Robinson A M, Young H K, Badofsky B, Rajguru S M, Stock S and Richter C P 2013 Behavioral and electrophysiological responses evoked by chronic infrared neural stimulation of the cochlea *PloS one* **8** e58189

[183] Duke A R, Cayce J M, Malphrus J D, Konrad P, Mahadevan-Jansen A and Jansen E D 2009 Combined optical and electrical stimulation of neural tissue in vivo *J Biomed Opt* **14** 060501

[184] Duke A R, Peterson E, Mackanos M A, Atkinson J, Tyler D and Jansen E D 2012 Hybrid electro-optical stimulation of the rat sciatic nerve induces force generation in the plantarflexor muscles *Journal of neural engineering* **9** 066006

[185] Cao Z, Xu Y, Suematsu N, Tan X, Young H and Richter C-P 2018 Hybrid Opto-electrical Neural Stimulation in Cochleae of Deaf White Cats. *Abstr. Assoc. Res. Otolaryngol.* **41** 607

[186] Bernstein R, Clough R L, Gillen K T, Malone M, Derzon D K, Griego A R, Tallant D, Garica M, Wheeler D and Harris D J Laboratory Assessment of Delaminated Polyimide Cable *Sandia National Laboratories* 1-12

[187] Kobayashi J, Matsuura T, Hida Y, Sasaki S and Maruno T 1998 Fluorinated Polyimide Waveguides with Low Polarization-Dependent Loss and Their Applications to Thermooptic Switches. *Journal of Lightwave Technology* **16** 1024-9

[188] Long E R and Long S A 1987 Spectroscopic Comparison of Effects of Electron Radiation on Mechanical Properties of Two Polyimides. *NASA* **NASA-TP-2663 19870009178**

[189] Okada T, Ishige R and Ando S 2016 Analysis of Thermal Radiation Properties of Polyimide and Polymeric Materials Based on ATR-IR spectroscopy. *Journal of Photopolymer Science and Technology* **29** 251-4

[190] Grossman N, Poher V, Grubb M S, Kennedy G T, Nikolic K, McGovern B, Berlinguer Palmini R, Gong Z, Drakakis E M, Neil M A, Dawson M D, Burrone J and Degenaar P 2010 Multi-site optical excitation using ChR2 and micro-LED array **7** 16004

[191] Poher V, Grossman N, Kennedy G T, Nikolic K, Zhang H X, Gong Z, Drakakis E M, Gu E, Dawson M D, French P M W, Degenaar P and Neil M A A 2008 Micro-LED arrays: a tool for two-dimensional neuron stimulation. *J. Phys. D: Appl. Phys.* **41** 094014 (9pp)

[192] Jiang H X and Lin J Y 2013 Nitride micro-LEDs and beyond--a decade progress review **21** A475-84

[193] Licklider J C R 1946 Effects of Amplitude Distortion upon the Intelligibility of Speech *The Journal of the Acoustical Society of America* **18** 249

[194] Smith Z M, Delgutte B and Oxenham A J 2002 Chimaeric sounds reveal dichotomies in auditory perception *Nature* **416** 87-90

[195] Lorenzi C, Gilbert G, Carn H, Garnier S and Moore B C 2006 Speech perception problems of the hearing impaired reflect inability to use temporal fine structure *Proceedings of the National Academy of Sciences of the United States of America* **103** 18866-9

[196] Plomp R 1983 *Hearing—physiological bases and psychophysics*: Springer) pp 270-6

[197] Burns E M and Viemeister N F 1981 Played-again SAM: Further observations on the pitch of amplitude-modulated noise *The Journal of the Acoustical Society of America* **70** 1655-60

[198] Lehiste I and Lass N J 1976 Suprasegmental features of speech *Contemporary issues in experimental phonetics* **225** 239

[199] 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 *The Journal of the Acoustical Society of America* **133** 2818-33

[200] Moon I J, Won J H, Park M H, Ives D T, Nie K, Heinz M G, Lorenzi C and Rubinstein J T 2014 Optimal combination of neural temporal envelope and fine structure cues to explain speech identification in background noise *The Journal of neuroscience : the official journal of the Society for Neuroscience* **34** 12145-54

[201] Leger A C, Desloge J G, Braida L D and Swaminathan J 2015 The role of recovered envelope cues in the identification of temporal-fine-structure speech for hearing-impaired listeners *The Journal of the Acoustical Society of America* **137** 505-8

[202] Leger A C, Reed C M, Desloge J G, Swaminathan J and Braida L D 2015 Consonant identification in noise using Hilbert-transform temporal fine-structure speech and recovered-envelope speech for listeners with normal and impaired hearing *The Journal of the Acoustical Society of America* **138** 389-403

[203] Moore B C and Glasberg B R 2010 The role of temporal fine structure in harmonic segregation through mistuning *The Journal of the Acoustical Society of America* **127** 5-8

[204] Moore B C 2014 Development and current status of the "Cambridge" loudness models *Trends Hear* **18**

[205] Dallos P and Cheatham M A 1976 Compound action potential (AP) tuning curves *The Journal of the Acoustical Society of America* **59** 591-7

[206] Eggermont J 1977 Compound actionpotential tuning curves in normal and pathological human ears *The Journal of the Acoustical Society of America* **62** 1247-51

[207] Fay R R 1988 *Hearing in vertebrates: a psychophysics databook*: Hill-Fay Associates Winnetka, IL)

[208] Evans E and Wilson J 1973 The frequency selectivity of the cochlea *Basic mechanisms in hearing* 519-54

[209] Zwicker E 1974 *Facts and models in hearing*: Springer) pp 132-41

[210] Shera C A, Guinan J J, Jr. and Oxenham A J 2002 Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements *Proceedings of the National Academy of Sciences of the United States of America* **99** 3318-23

[211] Dorman M F and Loizou P C 1998 The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels *Ear and hearing* **19** 162-6

[212] Whiteside S P 1998 Identification of a speaker's sex: a fricative study *Percept Mot Skills* **86** 587-91

[213] Whiteside S P 1998 Identification of a speaker's sex: a study of vowels *Percept Mot Skills* **86** 579-84

[214] Moore B C and Glasberg B R 1993 Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech *The Journal of the Acoustical Society of America* **94** 2050-62

[215] Moore B C, Glasberg B R and Vickers D A 1995 Simulation of the effects of loud ness recruitment on the intelligibility of speech in noise *British journal of audiology* **29** 131-43

[216] Moore B C, Wojtczak M and Vickers D A 1996 Effect of loudness recruitment on the perception of amplitude modulation *The Journal of the Acoustical Society of America* **100** 481-9

[217] Henry K S, Kale S and Heinz M G 2014 Noise-induced hearing loss increases the temporal precision of complex envelope coding by auditory-nerve fibers *Frontiers in systems neuroscience* **8** 20

[218] Henry K S, Kale S and Heinz M G 2016 Distorted Tonotopic Coding of Temporal Envelope and Fine Structure with Noise-Induced Hearing Loss *The Journal of neuroscience : the official journal of the Society for Neuroscience* **36** 2227-37

[219] Richter C P, Bayon R, Izzo A D, Otting M, Suh E, Goyal S, Hotaling J and Walsh J T, Jr. 2008 Optical stimulation of auditory neurons: effects of acute and chronic deafening *Hearing research* **242** 42-51

[220] Richter C P 2012 Optical Stimulation of the Auditory Nerve

[221] Chen M, Xu Y, Tan X and Richter C P 2016 Phase Processing in the Auditory System *Association for Research in Otolaryngology Abstract book* **39** 177

[222] Winslow R, Barta P and Sachs M 1987 Rate coding in the auditory nerve *Auditory processing of complex sounds* 212-24

[223] Conley R A and Keilson S E 1995 Rate representation and discriminability of second formant frequencies for/ɛ/-like steady-state vowels in cat auditory nerve *The Journal of the Acoustical Society of America* **98** 3223-34

[224] Richter C, Rajguru S, Matic A, Moreno E, Fishman A, Robinson A, Suh E and Walsh Jr J 2011 Spread of cochlear excitation during stimulation with pulsed infrared radiation: inferior colliculus measurements *Journal of neural engineering* **8** 056006

[225] Richter C-P and Matic A I 2011 *Auditory Prostheses*: Springer) pp 135-56

[226] Slaney M 1998 Auditory toolbox *Interval Research Corporation, Tech. Rep* **10** 1998

[227] Moore B C, Peters R W and Glasberg B R 1996 Detection of decrements and increments in sinusoids at high overall levels *The Journal of the Acoustical Society of America* **99** 3669-77

[228] McKay C M and McDermott H J 1998 Loudness perception with pulsatile electrical stimulation: the effect of interpulse intervals *The Journal of the Acoustical Society of America* **104** 1061-74

[229] Forrest T and Green D M 1987 Detection of partially filled gaps in noise and the temporal modulation transfer function *The Journal of the Acoustical Society of America* **82** 1933-43

[230] Lim H H, Lenarz M and Lenarz T 2009 Auditory midbrain implant: a review *Trends Amplif* **13** 149-80

[231] Langner G and Schreiner C 1987 Topology of functional parameters in the inferior colliculus of the cat *New frontiers in brain research* 122

[232] Krishna B S and Semple M N 2000 Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus *Journal of neurophysiology* **84** 255-73

[233] Seshagiri C V and Delgutte B 2007 Response properties of neighboring neurons in the auditory midbrain for pure-tone stimulation: a tetrode study *Journal of neurophysiology* **98** 2058-73

[234] Bader R 2015 Phase synchronization in the cochlea at transition from mechanical waves to electrical spikes *Chaos* **25** 103124

[235] Moller A R 1983 Frequency selectivity of phase-locking of complex sounds in the auditory nerve of the rat *Hearing research* **11** 267-84

[236] Sachs M B 1984 Neural coding of complex sounds: speech *Annual review of physiology* **46** 261-73

[237] Shamma S A 1985 Speech processing in the auditory system. I: The representation of speech sounds in the responses of the auditory nerve *The Journal of the Acoustical Society of America* **78** 1612-21

[238] Ehret G and Merzenich M M 1988 Complex sound analysis (frequency resolution, filtering and spectral integration) by single units of the inferior colliculus of the cat *Brain research* **472** 139-63

[239] Perez C A, Engineer C T, Jakkamsetti V, Carraway R S, Perry M S and Kilgard M P 2013 Different timescales for the neural coding of consonant and vowel sounds *Cerebral cortex* **23** 670-83

[240] Yates G K, Winter I M and Robertson D 1990 Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range *Hearing research* **45** 203-19

[241] Woolf N K, Ryan A F and Bone R C 1981 Neural phase-locking properties in the absence of cochlear outer hair cells *Hearing research* **4** 335-46

[242] Dynes S B and Delgutte B 1992 Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea *Hearing research* **58** 79-90

[243] Burger R M and Pollak G D 1998 Analysis of the role of inhibition in shaping responses to sinusoidally amplitude-modulated signals in the inferior colliculus *Journal of neurophysiology* **80** 1686-701

[244] Smalt C J, Krishnan A, Bidelman G M, Ananthakrishnan S and Gandour J T 2012 Distortion products and their influence on representation of pitch-relevant information in the human brainstem for unresolved harmonic complex tones *Hearing research* **292** 26-34

[245] Kato T, Fujita K and Kashimori Y 2015 A neural mechanism of phase-locked responses to sinusoidally amplitude-modulated signals in the inferior colliculus *Bio Systems* **134** 24-36