

1 Introduction

The normal hearing (or auditory) process is a highly non-linear biological signal processing system. Younger adults with normal hearing (YNH) have a remarkable ability to understand the speech of a human speaker in the presence of multiple speakers. This ability to focus on a specific human voice in a complex listening environment is referred to as a 'cocktail party effect' (Cherry, 1953). However, this distinct ability to identify the speech of a target speaker is reduced due to changes in the speech characteristics (e.g., shorter duration and lower sound level), and physiological and anatomical changes resulting due to increase in age and hearing loss. Generally, a YNH listener utilizes different cues to identify the speech during a multi-speaker scenario. Some of these cues are the differences in sound levels and their durations, the variations in speech spectral characteristics (e.g., formant difference cues), and the difference in fundamental frequency (F0) between the speakers (Bregman, 1990; Brokx and Nootboom, 1982; Calandruccio et al., 2019; Cherry, 1953; Han et al., 2021; Meister et al., 2020; Zwicker, 1984). Among these various cues, F0 difference is widely studied as it is commonly used to understand the segregation of speech sounds (Micheyl et al., 2010).

Concurrent vowel identification is a classical experimental paradigm to understand the role of F0 difference in the perceptual segregation of two synthetic vowels (simplest form of speech). In this experimental paradigm, two steady-state synthetic vowels with equal duration and sound level are presented simultaneously to one ear (i.e., monaural presentation) of a listener. The task of

the listener is to identify the two vowels by varying the F0 difference between them. This controlled experimental setup had enabled the researchers to study the effect of F0 difference as a segregation cue on vowel identification. Figure 1.1 shows the concurrent vowel scores for YNH subjects. The common observation is that the percent identification of both vowels improves with increasing F0 difference between the two vowels and then typically asymptotes at ~3 Hz F0 difference or higher (Arehart et al., 1997, 2005; Assmann and Summerfield, 1990; Chintanpalli et al., 2016; Chintanpalli and Heinz, 2013; Culling and Darwin, 1993; Summerfield and Assmann, 1991; Summers and Leek, 1998; Vongpaisal and Pichora-Fuller, 2007).

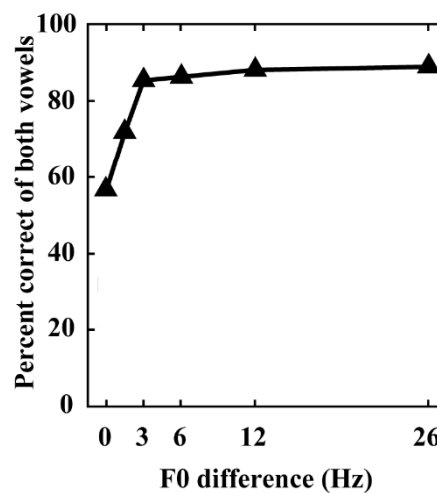


Figure 1-1 Effect of F0 difference on concurrent vowel scores (Chintanpalli et al., 2016).

The speech recognition ability in a multi-speaker scenario depends on the speaker's speech characteristics and the listener's hearing ability. The duration and sound level of the speech can limit the neural mechanism underlying the identification for YNH listeners. Apart from these acoustical changes, the neural mechanism can also be limited due to the auditory changes due to increases in age or hearing loss. Hence, there is a need to understand the neural

mechanisms underlying identification with changes in (1) acoustic and (2) physiological and anatomical factors resulting from increase in age and hearing loss. These issues had motivated the researchers to conduct various behavioral studies to understand the concurrent vowel scores across F0 differences for (1) YNH listeners with the acoustic variations (e.g., vowel duration and sound level) and (2) normal-aging and hearing impaired listeners. For YNH subjects, the concurrent vowel scores across F0 differences are reduced with the shorter duration vowel, when compared with the longer duration vowel (Assmann and Summerfield, 1990, 1994; Culling and Darwin, 1993; McKeown and Patterson, 1995). The level-dependent changes on concurrent vowel scores are increased from low to mid-levels and declined at higher levels for same and different F0 conditions in the YNH subjects (Chintanpalli et al., 2014). With increased age and hearing loss, the concurrent vowel identification scores across F0 differences are also reduced across F0 differences (Arehart et al., 1997, 2005, 2011; Chintanpalli et al., 2016; Snyder and Alain, 2005; Summers and Leek, 1998; Vongpaisal and Pichora-Fuller, 2007). These findings suggest there is a need to develop a biological-based signal processing model to understand the possible neural mechanisms underlying the identification due to acoustic and auditory changes across listeners.

There are computational models that validate the effect of F0 difference on concurrent vowel identification for a particular duration (e.g., 400 ms) and sound level (e.g., 65 dB sound pressure level (SPL)) for the YNH subjects (Chintanpalli and Heinz, 2013; Meddis and Hewitt, 1992). However, there are no modeling studies in the literature that could validate and explain the concurrent vowel identification scores across different sound levels and shorter durations

for YNH subjects. Additionally, there are also no modeling studies that can analyze the effects of age and hearing loss on concurrent vowel scores across F0 differences. These research gaps had motivated the current dissertation work to understand the possible neural mechanisms for these reduced concurrent vowel scores in YNH (duration and level), normal-aging subjects and hearing-impaired subjects with aging. More specifically, in this dissertation work, a physiologically based computational model will be developed to predict the concurrent vowel scores for: (1) different levels in the YNH subjects, (2) shorter duration in the YNH subjects and (3) normal-aging subjects and hearing-impaired subjects with aging. The concurrent vowel data from different behavioral studies will be used to validate the model predictions. The concurrent vowel modeling predictions in this research work will provide the underlying neural mechanisms for the challenges faced by the listeners due to acoustic (duration and sound level) and auditory (age and hearing loss) changes in a complex listening environment.

1.1 Background: Auditory System

The auditory system is responsible for hearing and is generally divided into peripheral and central auditory systems. The speech signal has to pass through both of these systems for its perception. Both systems play a crucial role in the perception of sounds by the listeners. However, this dissertation primarily investigates the effect of peripheral system on speech perception; hence it will be described in a greater depth.

1.1.1 Ear: Anatomy and Physiology

Figure 1-2(A) shows the diagram of the human ear. The peripheral auditory system includes the outer ear, middle ear, inner ear and auditory-nerve (AN) fibers, whereas the central auditory system includes many subsystems ranging from the cochlear nucleus to primary auditory cortex (located in the brain). An ear converts any acoustic signals [e.g., sinusoidal tone in Fig. 1-2(A)] into neural spikes at the AN fiber. The sound waves enter the outer ear and then pass through the external auditory canal, which makes the eardrum (or tympanic membrane) to vibrate [Fig. 1-2(A)]. These vibrations are sent to the middle ear. The middle ear transmits these vibrations to the inner ear with minimal energy loss due to impedance matching [Fig. 1-2(A)]. The inner ear contains a fluid-filled coil-shaped structure called the cochlea, which converts the vibrations from the middle ear into neural spikes at the AN fiber [Fig. 1-2(A)]. Inside the cochlea, a basilar membrane (BM) and an organ of corti [Fig. 1-2(B)] runs along the length of the cochlea to encode the various frequencies of the speech signal. Each place in the BM is tuned to a particular frequency. More specifically, the higher frequencies are encoded at the basal range of the cochlea, whereas the lower frequencies are encoded at the apical range of the cochlea. In other words, BM acts like a spectrum analyzer in terms of signal processing. The organ of corti contains inner-hair-cells [IHCs, Fig. 1-2(B)] and outer-hair-cells [OHCs, Fig.1-2(B)], which are responsible for normal hearing at the periphery.

The OHCs [green color in Fig. 1-2(B)] either amplify or reduce the BM vibrations as per the input level of the acoustic signal. If the sound level is very high, then the OHC reduces the BM vibrations to protect the ear from hearing

loss. Similarly, if the sound level is too soft, then the OHC amplifies the BM vibrations so that soft level sounds can be heard. The OHCs are responsible for cochlear non-linearities that are important for the normal functioning of cochlea. These properties are BM compression, two-tone suppression and distortion

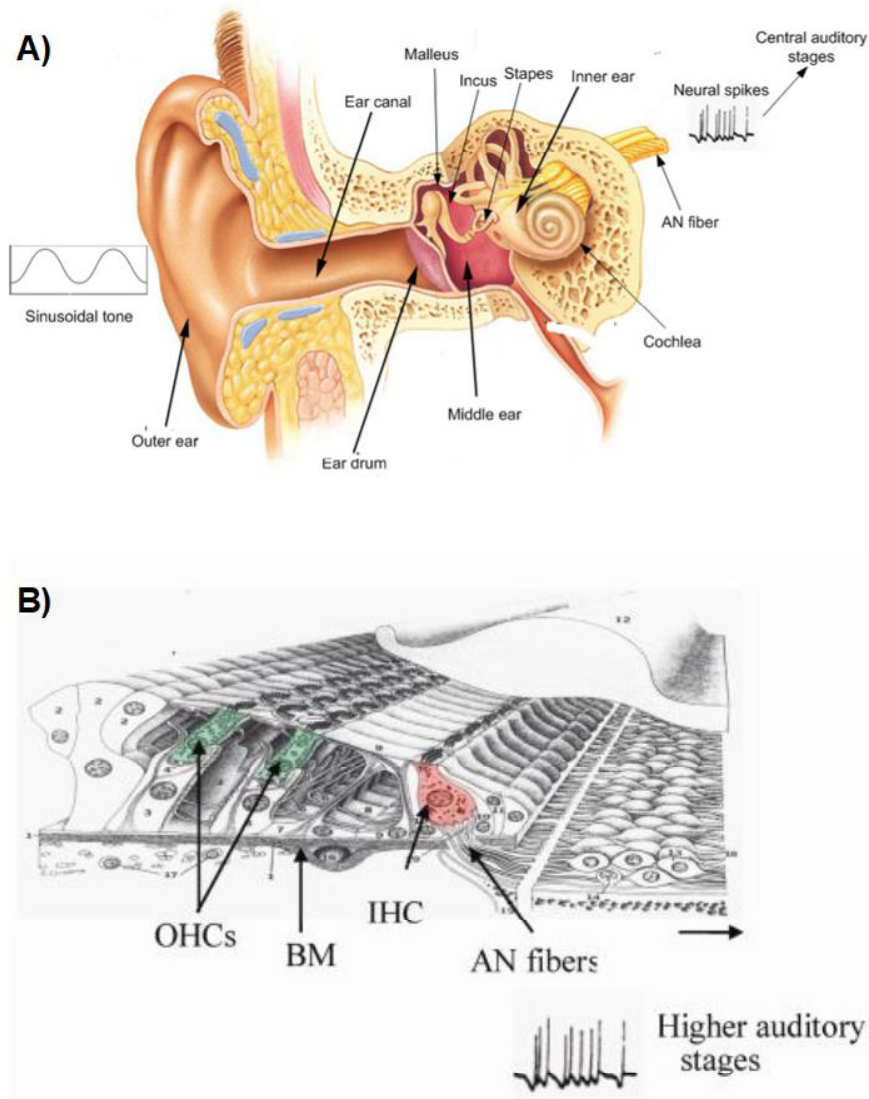


Figure 1-2 A) Anatomy of the human ear (<https://www.highpointaudiological.com/hearing-loss>). B) Major components of the cochlea (Junqueira et al., 1977). The input to the normal ear can be any acoustic signal and the output is the neural spikes of the corresponding acoustic signal.

product otoacoustic emissions (DPOAE). For each position (or frequency) along the length of cochlea, the BM vibration would be maximum to a particular stimulus frequency called as the characteristic frequency (CF). The BM vibration (in terms of velocity (mm/s)) shows a linear response at low levels, compressive

response at moderate levels and again a linear response at higher levels (Charaziak et al., 2020; He et al., 2021; Rhode, 1971, 1978; Sellick et al., 1982). In two-tone suppression, the BM response to a particular CF tone is reduced due to the presence of louder level at non-CF tone (Patuzzi et al., 1984; Rasetshwane et al., 2019; Siegel et al., 1982; Zhou and Nam, 2019). DPOAEs are cochlea generated sounds that are measured at the ear canal. When two different tones (one at frequency f_1 and another at f_2) are presented to the normal ear, DPOAEs ($mf_1 \pm nf_2$, where m and n are integers) are generated and the most dominant one occurs at $2f_2 - f_1$ (Liberman et al., 1996; Siegel et al., 1982; Sisto et al., 2018; Zelle et al., 2020). These non-linearities are reflected at the level of AN fibers and are described in the next section.

The IHCs [red color in Fig. 1-2(B)] translate the BM vibrations, modified by OHCs, into neural spikes at the AN fiber [Fig. 1-2(B)]. These spikes are later transmitted through the multiple stages of the central auditory system for sound perception. A hearing loss due to a dysfunction of the outer or middle ear is called as conductive hearing loss, while any damage to the hair cells (OHCs or IHCs) or loss of AN fibers results in a sensorineural hearing loss (SNHL). Conductive hearing loss is often corrected with hearing aids or through minor surgery. However, SNHL is irreversible because of the reduced cochlear non-linearities due to damages in hair cells and AN fibers. The most common form of SNHL is noise-induced hearing loss (~2/3 OHC damage and 1/3 IHC damage, Plack et al., 2004). The AN fibers send these neural spikes of the acoustic sound from cochlea to the central auditory system. There are around 50,000 fibers in cats and 30,000 fibers in humans. These fibers span the entire length of BM. Many single-unit recordings across various animal subjects like cats, chinchillas,

guinea pigs and gerbils are obtained to simple and complex stimuli are used to understand AN physiology (e.g., Anderson et al., 1971; Davis et al., 2007; Evans and Palmer, 1980; Hauser et al., 2018; Kajikawa et al., 2011; Li et al., 2015; Liberman, 1978; Liberman and Dodds, 1984a; Ngan and May, 2001; Sachs and Abbas, 1974; Trevino et al., 2019; Winter et al., 1990; Young and Barta, 1986).

1.1.2 Properties (Physiology) of auditory-nerve fibers

The AN fiber properties are obtained using the neural responses to simple stimuli (e.g., pure tones) and are described below.

1.1.2.1 Tuning-curve

The number of AN spikes produced per second in the absence of acoustic stimulation is called as the spontaneous rate (SR). During an acoustic stimulation, the number of AN spikes produced per second above SR is defined as driven rate. The driven and spontaneous rates together is referred as a neural discharge rate (DR). Tuning curve is a plot of frequency versus the lowest sound level for a fiber. The lowest sound level for each tone frequency is obtained if DR is greater than 10 spikes/sec above SR (Liberman, 1978) [dotted lines in Fig. 1-3(A) and 1-3(B)]. The parameters that are found from the tuning curve are CF, fiber's threshold and Q_{10} value. The frequency at which fiber responds to the lowest level is called CF. The threshold value is the sound level at CF. The Q_{10} is a measure of tuning and is the ratio of CF to the 10-dB bandwidth above the threshold. The larger Q_{10} value implies a sharper bandwidth (or sharp tuning) and the smaller Q_{10} value implies a broader bandwidth. The normal tuning curve has a high auditory sensitivity (i.e., lower threshold value) and sharp tuning [dotted lines in Fig. 1-3(A) and 1-3(B)] (Kiang et al., 1976; Liberman and Dodds,

1984b). The high sensitivity helps to detect the soft sound and sharp tuning helps to discriminate two or multiple frequency components of the speech signals. The IHC damage results in the loss of sensitivity without affecting the tuning [dashed line in Fig. 1-3(A)] (Liberman, 1978), whereas an OHC damage results in the loss of sensitivity and broad tuning [dashed line Fig. 1-3(B)] (Liberman and Dodds, 1984b). If the listener is having only an IHC damage, amplification can compensate for the loss of auditory sensitivity (Liberman and Dodds, 1984b). If the listener is having only an OHC damage, amplification can only compensate for the loss of auditory sensitivity, but the tuning cannot be sharpened as it will remain broaden [dashed line Fig. 1-3(B)] (Kiang et al., 1976; Liberman and Dodds, 1984b; Robertson, 1982). As mentioned previously, SNHL is a mixture of both OHC and IHC damages. Thus, it becomes very difficult to restore normal hearing because the current hearing aids (provides non-linear amplification with frequency) cannot solve the broad tuning problem caused due to OHC damage. Thus, it is difficult to develop biological-based signal processing algorithms to restore the normal tuning, which is currently a challenge for hearing-aid companies.

1.1.2.2 Rate-level function

Rate-level function provides fiber's DR (spikes/second) across sound levels. Figure 1-3(C) shows the rate-level function of a sinusoidal tone at CF from a normal AN fiber. As the level increases, DR increases above the SR and eventually saturates at higher levels [Fig. 1-3(C)]. The parameters derived from rate-level function are the threshold, saturation and dynamic range [Fig. 1-3(C)]. Threshold is the minimum sound level, where the fiber's DR increases 10 spikes/sec above the SR. The fiber reaches its saturation at higher levels, where

there is no further increase in the DR. The dynamic range is the difference in level at which the fiber saturates and threshold level. Based on the number of neural spikes per second, the AN fiber is either divided into high-SR (HSR, $SR > 18$ spikes/sec), medium-SR (MSR, $0.5 \leq SR < 18$ spikes/sec) or low-SR (LSR, $SR < 0.5$ spikes/sec) (e.g., Liberman, 1978; Winter et al., 1990). For cat's AN fibers, the HSR fibers comprise approximately 61%, the MSR fibers comprise 23% and the remaining 16% of the population are the LSR fibers (Liberman, 1978). The rate-level function parameters can be used to identify the AN fiber type (i.e., HSR, MSR and LSR). More specifically, the HSR fiber will have a lower threshold, narrower dynamic range and saturate at lower level, compared to the MSR and LSR fibers (Evans and Palmer, 1980; Sachs and Abbas, 1974).

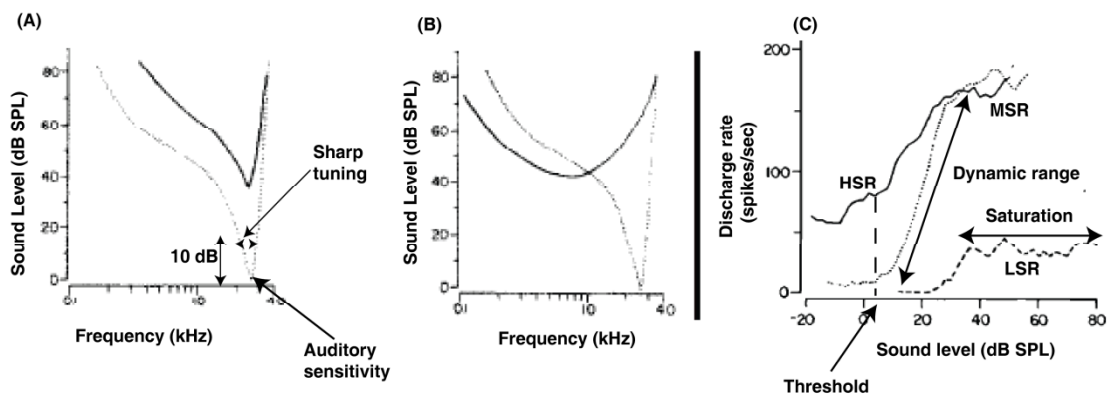


Figure 1-3 Cat tuning-curves and rate-level functions. Tuning curves for normal fiber (dotted lines). (A) IHC loss (dashed line) (B) OHC loss (dashed line) (both adapted from (Liberman and Dodds, 1984b)). (C) Rate-level functions for three fibers (Liberman, 1978). Threshold is the minimum sound level where DR is above SR, the fiber saturates at higher sound levels and the dynamic range specifies the limits where there are variations in DR. Notice that the LSR fiber have a high threshold, larger DR and saturate at higher level, as compared with the MSR and the HSR fibers.

1.1.2.3 Phase-locking

Phase-locking is defined as the ability of the AN spikes to lock to the stimulus frequency (Bidelman and Powers, 2018; Rose et al., 1967; Verschouten et al., 2019; Xu et al., 2017). The period histogram is used to characterize the

phase-locking of AN fiber and is obtained by counting the number of AN spikes produced over one period of the stimulus frequency for multiple representations of the stimulus (Fig. 1-4). The occurrence of these AN spikes is random; however, these spikes are largely produced at the positive half-cycle stimulus, across all the sound levels (Fig. 1-4). The phase-locking in the AN fibers is observed up to 5 kHz (especially in the cat, Miller et al., 1997). Eventually, the ability to phase-lock progressively diminishes with an increase in frequency with a relatively low strength still observed at 10 kHz (Anderson et al., 1971; Delgutte and Kiang, 1984; Miller et al., 1997; Rose et al., 1967; Sachs and Kiang, 1968).

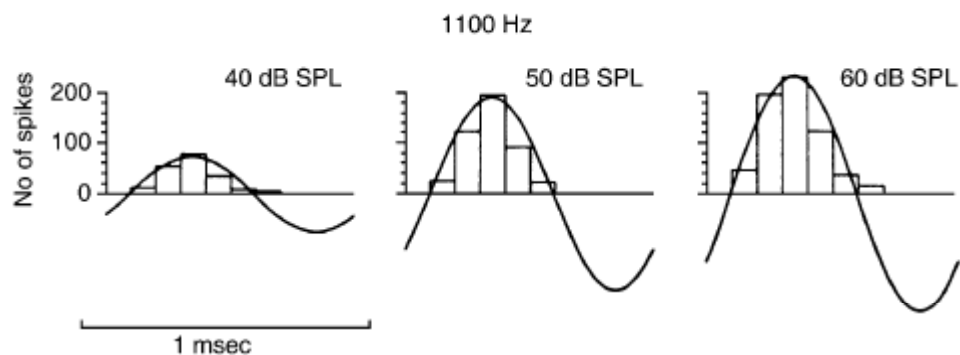


Figure 1-4 Period histograms of a fiber at different levels (Rose et al., 1971). Period histogram is obtained over one period of stimulus frequency (= 1100 Hz). For all the levels, the spikes are evoked largely in one-half of the cycle but with a fixed phase of frequency tone. The phase-locking is robust across sound levels.

1.1.2.4 Adaptation

Peri-stimulus time histogram is used to characterize the adaptation property of AN fiber, which is obtained by counting the number of AN spikes produced for a given time interval for multiple presentations of the stimulus (Rhode and Smith, 1985; Taberner and Liberman, 2005). The fiber's DR is high at the stimulus onset and then gradually adapts to the stimulus over its duration (Fig. 1-5). The adaptation property is mainly dependent on the SR of the fiber

(Müller and Robertson, 1991; Relkin and Doucet, 1991; Rhode and Smith, 1985). The HSR fibers have a higher onset response and adapt quickly to the stimulus when compared with the MSR and LSR fibers (Relkin and Doucet, 1991).

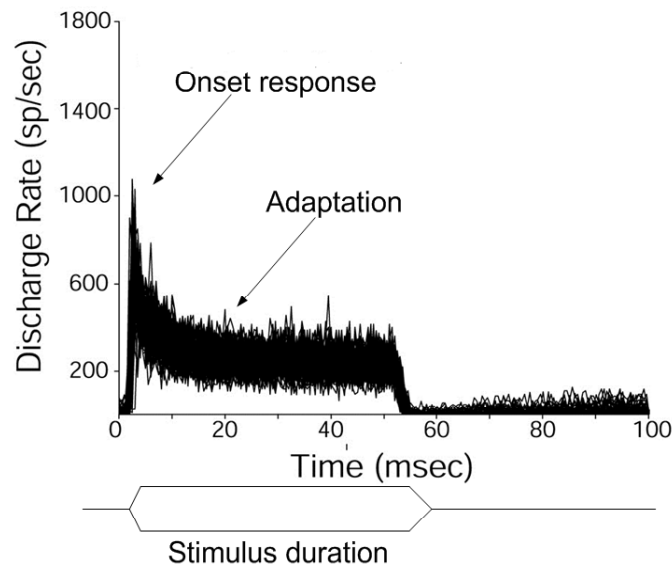


Figure 1-5 Peri-stimulus time histogram of AN fiber, over the stimulus duration. In this case, DR shows high onset response and adapts quickly to the stimulus (Taberner and Liberman, 2005).

1.1.2.5 Other physiological AN properties

As the stimulus level is increased, fiber tuning is decreased (or bandwidth is increased) along with a shift in CF. This property is referred to as a level-dependent tuning and CF-shift (Carney, 1999; Carney et al., 1999; Rose et al., 1971). For CF > 1.5 kHz, the shift is towards the lower value, whereas, for CF < 0.75 kHz the shift is towards the higher value. There is no shift in the intermediate region (Carney, 1999; Carney et al., 1999; Cheatham and Dallos, 2001; Rose et al., 1971). One of the important non-monotonic behaviors in AN responses to tones at high levels is the component 1- component 2 (C1/C2) transition (Gifford and Guinan, 1983; Heinz and Young, 2004; Kiang and Moxon, 1972; Wong et al., 1998). The C1/C2 transition occurs at a level, when there is a sudden change

in the phase of 180° with increasing sound level (Lieberman and Kiang, 1984). Additionally, this transition level can also be obtained by a decrease in rate-level function after saturation (often called as Nelson notch). For higher sound levels (> 90 dB SPL), the responses after the phase shift is referred as the C2 response, whereas the responses before the phase shift is called as the C1 response (sound levels < 90 dB SPL).

1.1.3 Auditory-nerve fiber representation for speech-like stimulus

Rate-place and temporal-place coding have been used to encode vowel features in AN fiber. The vowel spectrum of /ε/ [Fig. 1-6(A)] is represented using the population responses of AN fibers, as described below:

1.1.3.1 Rate-place coding

In rate-place coding, fiber's DR is plotted against the CF at a given sound level. The rate-level function, as specified above, is used for this coding scheme. Figure 1-6(B) and 1-6(C) shows the rate-level coding for a given vowel /ε/ at two levels (28 and 58 dB) (Sachs and Young, 1979). At lower sound levels (e.g., 28 dB), the local maxima of the DR for HSR fiber represents formants [as shown by downward arrows in Fig. 1-6(B)]. However, the same formants disappear at higher sound levels [solid line in Fig. 1-6(C)] due to fiber's rate saturation. An alternative solution could be to use the LSR fibers at higher sound levels because of their higher dynamic range in the rate-level function (Delgutte and Kiang, 1984; Sachs and Young, 1979). However, there are only 16% of LSR fibers available (out of 50,000 fibers in cats). Eventually, these fibers may also deteriorate because of the saturation rate (Sachs and Young, 1979).

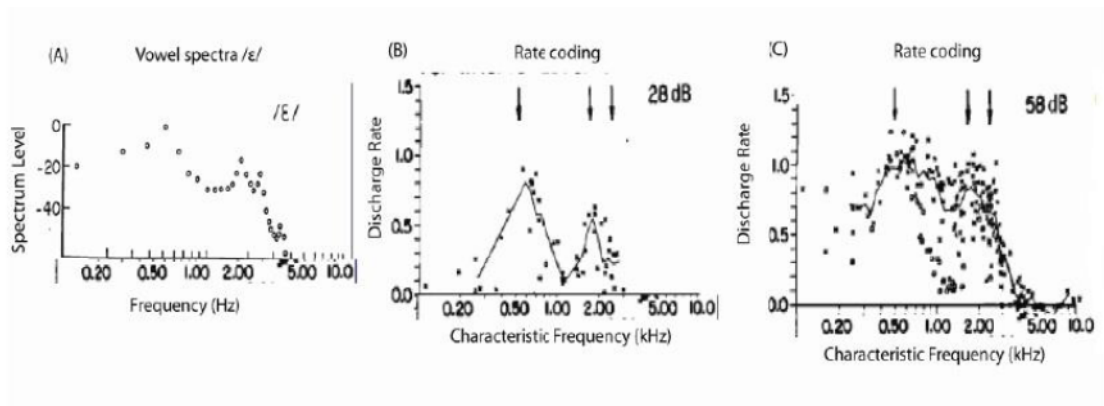


Figure 1-6 A) Frequency spectrum of vowel /ε/. Note that y-axis is in dB, (B – C) Rate coding of the vowel /ε/ at 28 dB and 58 dB. The downward arrows in (B) and (C) correspond to formants (F1, F2, F3) of /ε/. Due to rate saturation of HSR fiber, formant peaks disappear at 58 dB (Sachs and Young, 1979).

1.1.3.2 Temporal-place coding

A temporal coding scheme is based on the phase-locking as discussed before and is quantified using the average localized synchronized rate (ALSR). It provides a neural spectrum of the acoustic vowel across wide range of sound levels (Fig. 1-7) and hence is more robust than rate-place coding (Young and Sachs, 1979). The ALSR at a given frequency (e.g., 1000 Hz) is the average value of the synchronized rates obtained from all the fibers within ± 0.25 octaves of that frequency. This procedure is repeated across many frequencies. The synchronized rate is the magnitude of the Fourier transform of PSTH (discussed in section 1.1.2.4) and indicates to which vowel formant a given fiber is responding (Miller et al., 1997; Young and Sachs, 1979).

Temporal-place coding has been studied extensively in vowels for the normal and impaired fibers (Bruce et al., 2003; Delgutte, 1980; Miller et al., 1997; Young and Sachs, 1979). The normal fiber is dominated by formant nearest to a fiber's CF [Fig. 1-8(A)], whereas the impaired fiber (caused due to acoustic trauma) responds to lower formants below the fiber's CF due to broader tuning [Fig. 1-8(B)].

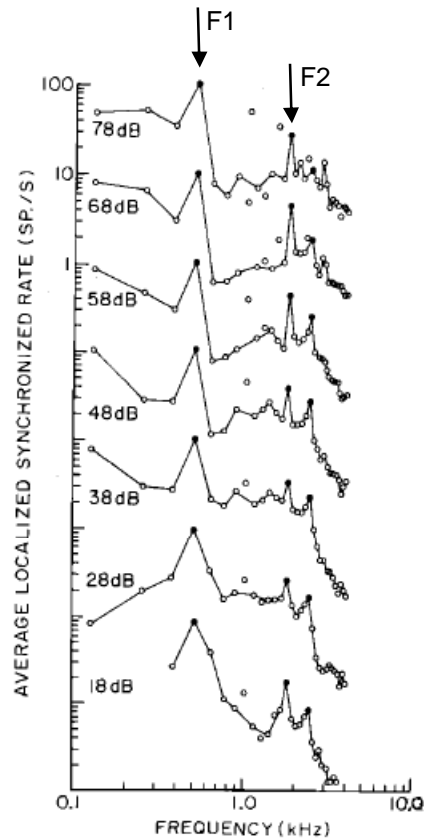


Figure 1-7 Average localized synchronized rate of the vowel /ε/ at various sound levels, ranging from 18 to 78 dB. Unlike rate coding, the formant peaks are well represented across levels. The downward arrows represent the locations of the formants (Young and Sachs, 1979).

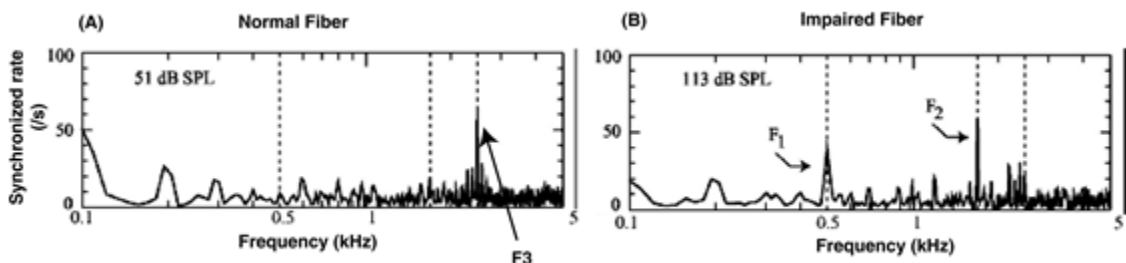


Figure 1-8 Synchronized rates in response to vowel /ε/ for two cat fibers with CF (≈ 2.5 kHz) near vowel formant F3. (A) Normal fiber. (B) Impaired fiber. Normal fiber responds to F3, whereas impaired fiber responds to a wide range of frequencies (including the lower formants F1 and F2). Notice that the impaired fiber is presented at a higher sound level to compensate for threshold shift caused by hearing loss (Miller et al., 1997).

1.1.4 Motivation behind developing signal processing based computational models

In neurophysiological experiments, data collection for a larger set of stimuli from animal subjects could be very challenging because of the time, cost and life of animals involved. Additionally, these experiments involve highly

intensive and expensive surgical procedures. To reduce the above-mentioned challenges in neurophysiological experiments (Holmes et al., 2004; Sachs et al., 2002), researchers are continuously developing a signal-processing based computational models that could qualitatively capture the AN physiological properties, as observed in animals. The major advantage of using the model is to obtain AN population responses (either through rate or temporal coding) that are required to predict the human identification scores. This dissertation work utilized the AN model approach to predict the concurrent vowel scores for YNH (across different levels and durations), normal-aging subjects and hearing-impaired subjects with aging.

1.1.4.1 Cat auditory-nerve model

Carney (1993) developed a computational model for the first time to predict AN responses to tones for HSR fibers at low CFs. The input stage of this model was a narrow-band filter that simulate the BM vibrations. The IHC transduction properties were simulated using a non-linear saturating function with two low pass filters, (having the same cut-off frequency = 1100 Hz). Some of the nonlinear properties (i.e., compressive nonlinearity, saturation and adaptation) were also incorporated. Zhang et al. (2001) further extended the model by including two-tone suppression by increasing the bandwidth of the control path filter bandwidth, relative to the signal path filter.

Bruce et al. (2003) further extended the Zhang et al. (2001) model to simulate impaired (and normal) AN responses. In order to simulate the impairment due to OHC and IHC loss, two model parameters C_{ohc} and C_{ihc} were introduced. The C_{ohc} and C_{ihc} values are ranged between 0 and 1; the lower the

value, the greater is the impairment. Apart from these, middle ear filter was also included in the AN model.

Zilany and Bruce (2006) extended the Bruce et al. (2003) model to simulate the AN responses at higher sound levels. This was achieved by including the effect of C1/C2 transition using two parallel paths. Their responses were then added and passed through the IHC model to produce the AN spikes. The first path was the output of C1 response that included the primary ('active') mode of the BM vibration, while the second path was the output of C2 response produced by a larger bandwidth filter for the simulation of broader tuning. For low and moderate sound levels (<90 dB SPL), the C1 output dominates, whereas the C2 output dominates at higher sound levels (>90 dB SPL). Zilany et al. (2007) verified the above model by qualitatively matching the temporal responses of a vowel /ɛ/ with the measured cat data (Miller et al., 1997) for both normal and hearing-impaired AN fibers. Zilany et al. (2007) also modeled the synchronized rates plots for normal and impaired AN fibers as shown in Fig. 1.8. Zilany et al. (2009) further added a model of rate adaptation between the IHC and AN fibers to simulate physiologically realistic adaptation properties for HSR, MSR and LSR fibers. Zilany et al. (2014) corrected the responses of higher CF model fibers to low frequency tones that were erroneously higher than the responses of low CF model fibers. More recently, Bruce et al. (2018) exhibit an improvement in several measures of AN fiber spiking statistics (e.g., high firing rates, temporal precision) and predicted rate-level functions that matched better with the physiological data obtained from cats.

1.1.4.2 Guinea pig auditory-nerve model

Lopez-Poveda and Meddis (2001) simulated the characteristics of the BM vibrations of the cochlea using the dual resonance non-linear (DRNL) filter bank, which contains two independent linear and non-linear paths. This DRNL filter bank simulated the BM vibrations in response to sinusoidal tones. Sumner et al. (2002) developed an IHC transduction model to produce the responses of AN fibers for all three types of fibers but only at high CFs (> 8 kHz). Subsequently, Sumner et al. (2003b) cascaded the DRNL model with an IHC transduction model (Sumner et al., 2003a) to produce the AN responses across all CFs and fiber types.

Lopez-Najera et al. (2007) developed a triple-path nonlinear filter bank (TRNL). The third path is a linear, low-gain, all-pass filter. The TRNL included the same DRNL but with a new parallel path that included a linear, low-gain all-pass filter. This was effective only to simulate BM vibrations at higher levels. However, the main focus of this group is to predict the normal AN responses for understanding the behavioral measures of normal hearing listeners. As this dissertation also focuses on hearing loss, AN model developed from cat's data will be used.

1.1.4.3 Justification for using animal subjects to study the human ear

A debatable question is whether AN responses from either animals or model are appropriate for humans. The general properties of cochlea and AN fibers are similar across cat and human; however, they differ in the frequency range of hearing and the layout of frequencies along the length of the cochlea. The human hearing range extends from 20 Hz - 20 kHz in a 35 mm cochlea and

the cat hearing range extends from 100 Hz - 60 kHz in a 25 mm cochlea (Greenwood, 1990; Kieft et al., 2001; Recio et al., 2002). However, when the frequency range and cochlear length are normalized across different species (human, cats, guinea pigs and chinchillas), the resulting cochlear frequency-position function is exponential (Greenwood, 1990; Kieft et al., 2001) and is given by:

$$x = \frac{1}{a} \log_{10} \left(\frac{F}{A} + k \right) \quad 1-1$$

Where x is the distance along the cochlea, F is the frequency, and parameters a , A and k are constants and are functions of species. Comparisons among species are mainly done on the tuning. Ruggero and Temchin (2005) stated that human tuning is similar to cats and guinea pigs. Whereas the more recent studies (Joris et al., 2011; Shera et al., 2010) suggest that human tuning is at least twice as sharp compared to animals.

Based on these qualitative comparisons, cat fibers can be used to study the neural coding of perception in human listeners; however, potential species differences still remain. Multiple researchers have used the animal AN model to predict the speech recognition scores of human behavioral data (Chintanpalli and Heinz, 2013; Encina-Llamas et al., 2019; Hedrick et al., 2016; Hines and Harte, 2010, 2012; Parthasarathy et al., 2016).

1.2 General approach used in this dissertation

A physiologically based computational modeling is used in this dissertation work to predict the concurrent vowel scores for the YNH subjects (across different durations and levels), normal-aging subjects and hearing-

impaired subjects with aging. The published identification data from normal, aged and hearing impaired listeners will be used to validate the model scores. These model predictions help to understand the neural mechanisms underlying the changes in the concurrent vowel scores.

1.3 Research Objectives

. This dissertation work utilized the AN model approach to predict the concurrent vowel scores for YNH subjects (across different levels and durations), normal-aging subjects and hearing-impaired subjects with aging. The model predictions of the current dissertation were obtained through the MATLAB software.

The objectives of the thesis can be summarized as follows:

1. Computational model predictions of level-dependent changes on concurrent vowel identification scores.
2. Computational model predictions of shorter duration on concurrent vowel identification scores.
3. Computational model predictions of age and hearing loss on concurrent vowel identification scores.

1.4 Overview of this dissertation work

Chintanpalli et al. (2014) have used the percent correct identification to understand the level-dependent changes in concurrent vowel scores across two different F0 differences for YNH subjects. **Chapter 2** had predicted these level-dependent changes through a physiologically based AN computational model and the F0-guided segregation algorithm. This modeling study helps to

understand the effect of sound level on the ability to utilize F0 difference cue for concurrent vowel identification.

YNH listeners generally have more difficulty in pitch-perception and F0-discrimination as the stimulus duration is reduced (Gockel et al., 2007; Moore et al., 1986). Similarly, concurrent vowel behavioral studies have shown that the effect of F0 difference is reduced on concurrent vowel scores, when the vowel duration is decreased. Two studies have shown a reduced F0 benefit (Assmann and Summerfield, 1994; McKeown and Patterson, 1995), while others have shown no F0 benefit (Assmann and Summerfield, 1990; Culling and Darwin, 1993). These studies suggest across-subjects variability in utilizing the F0 difference cue for identification. **Chapter 3** investigates the predicted effects of shorter vowel duration on concurrent vowel scores using the AN model and the F0-guided segregation algorithm. Predictions based on the reduced F0 guided segregation cue was used to understand the effects of reduced concurrent vowel scores at a shorter vowel duration. Additionally, differential reductions in the ability to avail F0-guided segregation cue might explain the individual differences in F0-benefit across YNH subjects.

Relative to the YNH subjects, the ability to identify the speech of the target speaker is reduced with increase in age and hearing loss. To understand the effects of age and hearing loss, Chintanpalli et al. (2016) collected the concurrent-vowel data across F0 differences for (YNH), older adults with normal-hearing (ONH) and older adults with hearing loss (OHI). The overall identification scores across F0 differences were reduced for both ONH and OHI subjects, but the lowest scores were observed for OHI subjects. **Chapter 4** predicted the

concurrent-vowel scores for these three listening groups to understand the effects of age and hearing loss by developing a physiologically based computational model. The YNH model was developed using the AN model and the F0-guided segregation algorithm. The ONH and OHI models were developed based on the peripheral changes due to increase in age and hearing loss. All three models had the same F0-guided segregation algorithm. This was done primarily to understand how peripheral changes due to age and hearing loss could affect the concurrent vowel scores. The model predictions were reduced across F0 differences for the ONH model and were successful in capturing the pattern of identification scores of concurrent-vowel data. For the OHI model, predicted concurrent-vowel scores were lowest and captured the pattern of identification scores of concurrent-vowel data. The reduced vowel segregation in both the ONH and OHI models suggested a limited use of the F0-difference segregation cue due to increase in age and hearing loss.

Finally, **Chapter 5** concludes with an overall discussion, limitations and possible future work following the dissertation. A clinical application related with this work is also discussed. Finally, the dissertation concludes with the possible changes in neural mechanisms, which could affect the concurrent vowel scores due to acoustic and auditory variations.