4 Modeling the effects of age and hearing loss on concurrent vowel scores[5,6]

4.1 Introduction

The ability to identify a target speaker in a multi-talker scenario is reduced due to anatomical and physiological changes in the auditory system for aged listeners and hearing impaired listeners. To understand these age and hearing loss effects, there has been various behavioral studies conducted on concurrent vowel identification. More specifically, several concurrent vowel identification studies have considered how the F0-difference cue affects the identification scores due to increased age and hearing loss. The identification scores of both vowels across F0 differences are reduced with increasing age (Arehart et al., 2011; Snyder and Alain, 2005; Vongpaisal and Pichora-Fuller, 2007). Other studies show that identification scores across F0 differences are also reduced with hearing loss (Arehart et al., 1997, 2005; Summers and Leek, 1998). Chintanpalli et al. (2016) is the only behavioral data in the concurrent vowel literature that addressed the effects of age and hearing loss on percent identification scores in a single study. Chintanpalli et al. (2016) collected concurrent vowel data across three different listening groups: YNH, ONH and OHI. The overall identification scores of both vowels across F0 differences were reduced for ONH subjects, compared with the YNH subjects. The scores for OHI subjects were the lowest across F0 differences. Furthermore, it was concluded that the F0 difference cue was important for identifying the second vowel of the

---

[5] This work was published in the Journal of the Acoustical Society of America, vol. 150(5), 3581 – 3592, 2021. doi: 10.1121/10.0007046

[6] The extension of this work was published in the 8th International Symposium on Auditory and Audiological (ISAAR) in Denmark, 2021.

vowel pair, as the percent correct of one vowel was ≥ 95% at each F0 difference for three listening groups. The percent correct of one vowel was computed as the proportion of vowel pairs (out of 25) in which one vowel was correctly identified in each pair. The F0 benefit is defined as the difference in percent identification of both vowels between 26-Hz and 0-Hz F0 difference conditions and is a quantitative metric used in the concurrent vowel literature (Arehart et al., 1997, 2005; Assmann and Summerfield, 1990; Chintanpalli et al., 2014, 2016; Settibhaktini and Chintanpalli, 2018, 2020; Summers and Leek, 1998). The YNH and OHI subjects had the largest (32%) and lowest (6.8%) F0 benefits, respectively, while ONH subjects had an intermediate value of 22%.

The increase in identification of both vowels for the YNH subjects is due to an improvement in segregating two vowels using the F0 difference cue. At zero-F0 difference, only formant difference cues between the two vowels are available for identification. With increasing F0 difference, the ability to segregate vowels improves and thereby assists in identifying both vowels. There have been many attempts to develop a computational model for YNH subjects that could capture the pattern of identification scores of both vowels (Assmann and Summerfield, 1990; Meddis and Hewitt, 1992; Scheffers, 1983; Zwicker, 1984). These models require identifying either one or both F0s correctly for segregating two vowels prior to identification. Among these, Meddis and Hewitt (1992) model successfully captured the gradual increase in the identification scores of both vowels with increasing F0 difference (0 – 6 Hz) and then the scores asymptoted at 6-Hz F0 difference. This model only requires the correct estimation of one F0 of the two vowels for proper segregation and correct identification. Meddis and

Hewitt (1992) model suggests that the phase-locking of auditory-nerve (AN) fibers to the F0s of the two vowels improves with increasing F0 difference, resulting in enhanced F0-guided segregation which contributes to correct identification. Subsequently, Chintanpalli and Heinz (2013) and Settibhaktini and Chintanpalli (2020) successfully tested the same F0-guided segregation algorithm (Meddis and Hewitt, 1992) but with more recent AN models (Zilany et al., 2014; Zilany and Bruce, 2007) for identifying both vowels as a function of F0 difference. Chintanpalli and Heinz (2013) also showed a relationship between F0-guided segregation (as measured using the percent F0 segregation metric) and the identification score for both vowels. More specifically, the improvement in the identification score was attributed to enhancement in percent F0 segregation with increasing F0 difference (also consistent with Settibhaktini and Chintanpalli, 2020). However, there is no computational model in the literature that could capture the F0 difference effect on concurrent vowel scores due to age and hearing loss. Thus, a computational model needs to be developed by incorporating the anatomical and physiological changes known to occur due to age and hearing loss.

The endocochlear potential (EP) is the cochlear battery applied between the scala media and scala tympani regions of the cochlea. It is essential for the normal functioning of the mechanoelectrical transduction of the hair cells. The normal EP value is ~ 90 mV. However, with increasing age, this value is reduced to ~ 60 mV or even lower, an effect that has been observed across various animal studies (e.g., Mills et al., 2006; Schmiedt et al., 2002, 1996). The reduction in EP resulted in significant threshold elevations as well as alterations in the shape of AN-fiber tuning curves (e.g., larger effects at the tip, but also reduced sensitivity in the tail, Schmiedt et al., 1990; Sewell, 1984). The significant effects at the tip were also seen in basilar-membrane responses following furosemide, but not the tail effects; suggesting OHC dysfunction was primarily responsible for the frequency-specific tip effects, while IHC dysfunction contributed to both tail and tip effects in a frequency-independent manner (Ruggero and Rich, 1991). Additionally, thresholds of AN-fiber rate-level functions were also elevated (Hellstrom and Schmiedt, 1991). In addition to these changes, there was a reduced effect of cochlear amplification as measured by basilar-membrane responses (Ruggero and Rich, 1991), compound action potentials (Gleich et al., 2016; Mills et al., 2006; Schmiedt et al., 2002, 1996) and distortion product otoacoustic emissions (Wang et al., 2019). As a whole, these findings suggest that EP reduction affects the functionality of both OHCs and IHCs, since the EP is the battery for both types of hair cells. When the EP was reduced using furosemide, the neural thresholds (measured using the compound action potentials) of young gerbils were correlated well with the thresholds of quiet-aged gerbils (Mills et al., 2006; Schmiedt et al., 2002). Dubno et al. (2013) used the audiogram profiles from animal models to successfully classify the audiograms of aged humans and concluded that the EP is a primary factor for audiometric threshold shifts.

Noise-induced hearing loss (NIHL) occurs primarily because of mechanical damage to the OHCs and IHCs, both through stereocilia damage and hair-cell death (Liberman and Dodds, 1984a, 1984b). Due to the OHC damage, the strength of cochlear nonlinearities (e.g., compression, suppression, broadened tuning and best-frequency shifts) is reduced (Rhode, 1971; Ruggero et al., 1997). There is also reduced frequency selectivity and sensitivity of AN fibers (e.g., Liberman and Dodds, 1984a; Miller et al., 1997a).. Due to the IHC damage, the sensitivity of AN fibers is reduced without affecting the frequency selectivity (e.g., Liberman and Dodds, 1984a; Miller et al., 1997a). Computational AN-modeling work suggests that AN-fiber tuning curve effects following NIHL are well accounted for by roughly 2/3 OHC and 1/3 IHC damage (Bruce et al., 2003). Physiological studies suggest that EP reduction is not a major contributor for permanent threshold shifts in NIHL animals (Kujawa and Liberman, 2019; Liberman and Gao, 1995; Liberman and Mulroy, 1982; Wang et al., 2002). Additionally, recent temporal-bone studies suggest that in older human adults, EP reduction (i.e., metabolic hearing loss) is not a significant factor compared to hair-cell damage (i.e., sensory loss) (Wu et al., 2020a, 2020b). However, these anatomical studies were unable to make functional evaluations of the combined effects of EP reduction and hair-cell damage in OHI subjects, which is a difficult issue to address as both effects assist directly to OHC and IHC functionalities.

With increasing age, there is a proportional loss of MSR and LSR AN fibers for CFs ≥ 6 KHz (Schmiedt et al., 1996). For NIHL, there is a proportional loss of HSR AN fibers, but there is a relative increase in MSR and LSR fibers (e.g., Heinz and Young, 2004; Liberman and Dodds, 1984b). Since phase-locking to pure tones of normal AN fibers does not depend strongly on SR distributions (Johnson, 1980), phase-locking may not be altered with age or NIHL. In fact this is the case, as the phase-locking of AN fibers to pure tones in quiet is not altered with increased age (Heeringa et al., 2020) nor with hearing loss (Harrison and Evans, 1979; Henry and Heinz, 2012; Miller et al., 1997). These findings suggest that the phase locking ability of AN fibers for a synthetic vowel in quiet may not be altered with SR variations for increased age and hearing loss.

Animal studies have shown that both normal aging and NIHL can lead to the loss of synaptic connections innervating the hair-cells (normal-aging: Sergeyenko et al., 2013; Stamataki et al., 2006; noise-induced hearing loss: Furman et al., 2013; Kujawa and Liberman, 2009; Liberman and Kujawa, 2017; Valero et al., 2017). This is referred to as cochlear synaptopathy (CS) and is hidden from audiometric results (Eckert et al., 2021; Liberman and Kujawa, 2017). This hidden effect is thought to alter speech recognition scores without any change in the audiometric thresholds (Grant et al., 2020; Mepani et al., 2020). Wu et al. (2019) showed that a large number of AN fibers are disconnected from their corresponding hair-cells based on temporal-bone analyses from normal-aging humans. It was concluded that over 60% of synaptic connections were lost, when averaged across the audiometric frequencies, for humans over 50 years of age. Johannesen et al. (2019) showed that the slope of auditory brainstem response wave-I amplitude with level was reduced with increasing age in humans, suggesting the presence of CS. The above-mentioned animal and human studies suggest that the number of AN fibers available to process the neural responses in the central auditory system is limited due to loss of cochlear synaptic connections, which might affect speech identification scores for both ONH and OHI listeners.

The current study models the effect of F0 difference on identification scores of concurrent vowels for three different listening groups (YNH, ONH, and OHI subjects). The concurrent-vowel data collected by Chintanpalli et al. (2016)

is used to validate these model predictions. The configuration of mean audiometric thresholds of the OHI subjects of Chintanpalli et al. (2016), also shown in Table 4-1 here, fall under the combined metabolic (reduced EP) and sensory categories of classified audiogram shapes, having acoustic overexposure (Dubno et al., 2013). This result combined with the recent anatomical studies suggesting that older adults have significant hair-cell damage, relative to the EP reduction (Wu et al., 2020a, 2020b). Hence, it is reasonable to assume that the OHI subjects in Chintanpalli et al. (2016) had combined EP reduction and NIHL, which both affect the OHC and IHC functionalities. We hypothesized that (1) reduction in the EP and CS (i.e., loss of AN fibers) might contribute to reduced concurrent-vowel scores across F0 differences in ONH subjects, (2) significant reduction in the functionalities of OHCs and IHCs (due to combined EP drop and mechanical hair-cell damage), and CS might contribute to the lowest concurrent-vowel scores across F0 differences in OHI subjects, and (3) the CS might be larger for OHI subjects due to the combined effects of hearing loss and age. For the OHI model, the EP loss and hair-cell damage are captured in the same way since they both affect OHC and IHC function. However, it is largely driven by the hair-cell damage effect, as the effect of EP loss appears to be secondary in the presence of NIHL (Kujawa and Liberman, 2019; Liberman and Gao, 1995; Liberman and Mulroy, 1982; Wang et al., 2002; Wu et al., 2020a, 2020b).

To test these hypotheses, a computational model for concurrent-vowel scores was developed using the population responses from a recent version of a well-established AN model (Bruce et al., 2018), with a modified Meddis and Hewitt (1992) F0-guided segregation algorithm. A similar modeling framework has been used previously to successfully capture the F0 difference effect on concurrent-vowel identification for YNH subjects (Chintanpalli and Heinz, 2013; Meddis and Hewitt, 1992; Settibhaktini and Chintanpalli, 2018, 2020). The goal of the current study was to determine whether the peripheral changes due to increased age and hearing loss can explain the mechanisms underlying the reduced concurrent-vowel scores across F0 differences for ONH and OHI subjects.

4.2 Methods

4.2.1 Stimuli

. The concurrent vowel stimulus generation were similar to the second objective (section 3.2.1), except that the duration was 400 ms (consistent with Chintanpalli et al., 2016). Overall, 150 concurrent vowels (25 pairs x 6 F0 differences) were presented as an input to each of the listening models. Consistent with Chintanpalli et al (2016), the individual vowel was presented at 65 dB SPL for YNH and ONH models, whereas 85 dB SPL was used for the OHI model. However, the vowel pair level was ~3 dB higher than the individual vowels in each model.

4.2.2 General Modeling Framework

Figure 4-1 shows the block diagram for each of the listening models used in the current study for predicting the identification of both vowels across F0 differences. In this framework, a physiologically realistic auditory-nerve model (Bruce et al., 2018) was cascaded with a modified version of the F0-guided segregation algorithm (Meddis and Hewitt, 1992) to understand the effect of F0 difference on concurrent vowel scores [Fig. 4-1(A)].

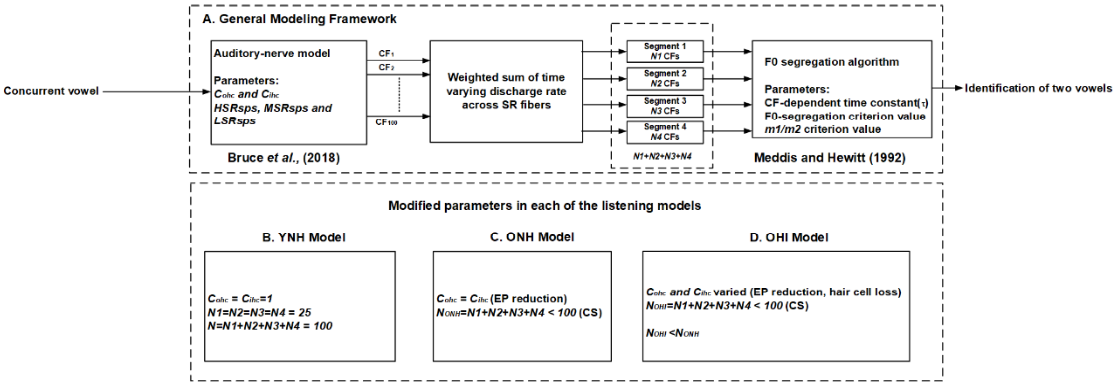

Figure 4-1 Block diagram illustrating the steps involved in the computational model to predict the concurrent vowel scores across F0 differences. (A) General modeling framework across three listening models. The parameters of AN model are C_{ohc}, C_{ihc}, *HSRsps*, *MSRsps*, and *LSRsps*, whereas the parameters for the F0-segregation algorithm are CF-dependent time constant ($\Delta\tau$), F0-segregation and *m1/m2* criterion values. The AN responses are obtained for 100 CFs that ranged logarithmically between 250 – 4000 Hz. These 100 CFs are divided into four octave-spaced segments [i.e., segment 1: (250 to 500 Hz), segment 2: (500 to 1000 Hz), segment 3: (1000 to 2000 Hz) and segment 4: (2000 to 4000 Hz)] using N_1, N_2, N_3, and N_4 for the low to high CF bands, respectively. (B) Parameters modified for YNH computational model. (C) Parameters modified for ONH computational model. (D) Parameters modified for OHI computational model. For the ONH and OHI models, the number of CFs in each segment is varied to simulate CS. For hypothesis testing, only the parameters associated with the peripheral stage were altered to predict the concurrent vowel scores for ONH and OHI models, whereas the same parameter values were used for the F0-segregation algorithm across all three peripheral models.

4.2.2.1 Auditory-nerve model

The third objective utilized the most recent version of the AN model (Bruce et al., 2018) to predict the temporal responses of concurrent vowels. This latest version of the model successfully simulates the temporal responses from different SR fibers, that is relevant to the current study. A key feature of this model is the ability to alter the functionalities of OHCs and IHCs, using the two distinct model parameters (i.e., C_{ohc} and C_{ihc}). These values range from 1 (normal function) to 0 (complete hair-cell loss). Intermediate values result in partial OHC or IHC damage. Relevant to the current study, the model captures (1) the level-dependent changes in cochlear nonlinearities for normal AN fibers, (2) the reduction in cochlear nonlinearities for impaired AN fibers, (3) the level-dependent changes in phase-locking to vowel formants and F0s (Chintanpalli et al., 2014; Settibhaktini and Chintanpalli, 2018; Zilany and Bruce, 2007) for normal AN fibers and (4) the reduction in phase-locking to vowel's higher formant frequencies (e.g., F2 and F3, due to enhanced F1 coding) for impaired AN fibers (Zilany and Bruce, 2007).

The AN model's input was a concurrent vowel and the output was a time-varying DR of AN fiber for a particular CF. These AN responses were predicted across 100 CFs, ranging from 250 to 4000 Hz. The DR was obtained at each of these CFs for different SR fibers. More specifically, as per the anatomical and physiological functions (Liberman, 1978), the AN population was classified into three SR distributions: HSR ($SR \geq 18$ spikes/sec), MSR ($0.5 \leq SR < 18$ spikes/sec) and LSR ($SR < 0.5$ spikes/sec).

To obtain the population of AN responses, each CF's overall DR was calculated as the weighted sum of the DRs as per the distribution of SR fibers (61% of HSR fibers, 23% of MSR fibers, and 16% of LSR fibers; Liberman, 1978). These weighted DRs across CFs were presented as inputs to a Meddis and Hewitt (1992) F0-guided segregation algorithm to predict concurrent vowel scores across F0 differences [Fig. 4-1(A)].

Table 4-1 Mean values of the pure-tone thresholds (dB HL) for YNH, ONH and OHI subjects which are obtained from the Fig. 1 of Chintanpalli et al. (2016).

	Frequency (Hz)							
Subjects	250	500	1000	2000	3000	4000	6000	8000
YNH	7.7	8.0	2.0	3.7	4.0	1.7	6.8	5.0
ONH	9.7	9.3	7.3	10.0	12.3	15.7	18.3	32.7
OHI	13.0	17.3	18.3	26.3	38.8	48.7	59	61

4.2.2.2 F0-guided segregation algorithm for concurrent vowel identification

The F0- guided segregation algorithm used in the third objective is the same that was used for first and second objectives (section 2.3.2.2 and Fig. 2-2). Only difference was that the CF-dependent time constant values were altered. In this objective, the time constant varied with CF ($\Delta\tau$ = 270 ms for 250 ≤ CF < 440 Hz; $\Delta\tau$ = 256 ms for 440 ≤ CF < 880 Hz; $\Delta\tau$ = 250 ms for 880 ≤ CF < 1320 Hz; $\Delta\tau$ = 249 ms for CF ≥ 1320 Hz). The range of time constant values across CFs were different compared to the first objective because of the different AN model used (Bruce et al., 2018). Here, we test this segregation algorithm's effectiveness to predict concurrent vowel scores for ONH and OHI models.

4.2.3 YNH model

The C_{ohc} and C_{ihc} values across selected CFs were equal to 1 to simulate the normal functionalities of OHCs and IHCs [Fig. 4-1(B)]. The parameters *HSRsps*, *MSRsps* and *LSRsps* (spikes/s) of Figure 4-1(A) corresponds to the SR values used for HSR, MSR and LSR fibers, respectively. These parameter values were chosen by matching the AN model's thresholds with the mean audiometric thresholds between 250 and 4000 Hz for YNH subjects. The upper limit was 4000 Hz, as all vowel formants are less than 4000 Hz (Table 2-1). Table 4-1 shows the mean audiometric thresholds for YNH, ONH and OHI subjects of Chintanpalli *et al.* (2016). At each audiometric frequency, the AN model's threshold was the minimum sound level required to produce a firing rate 10 spikes/s above SR, obtained using the peri-stimulus time histogram of AN population responses. If the relative pure-tone average (PTA) difference between the actual and model thresholds was less than 10%, then the thresholds were judged to match in the current study.

4.2.4 ONH model

The ONH model included the effects of EP loss and CS to predict the concurrent-vowel scores across F0 differences [Fig. 4-1(C)]. In the current study, the EP reduction was assumed to affect both types of hair cells to an equal extent (i.e., $C_{ohc} = C_{ihc}$). To simulate a reduction in the EP, the equal C_{ohc} and C_{ihc} values were varied to match the AN model's thresholds with the mean audiometric thresholds between 250 and 4000 Hz for ONH subjects (Table 4-1). The number of logarithmically spaced CFs in each octave segment were reduced using the parameters N_1, N_2, N_3 and N_4 of Figure 4-1(C) to approximate the reduced number of AN fibers sending information centrally due to CS. To predict the ONH model's identification scores, the AN responses obtained using the $C_{ohc} = C_{ihc}$ (Bruce et al., 2018), were passed through a modified F0-guided segregation algorithm (Meddis and Hewitt, 1992) with the reduced number of CFs (for computing the ACFs) to predict the concurrent vowel scores.

4.2.5 OHI model

The effects of hair-cell loss (and EP reduction) and CS were included in the OHI model to predict the concurrent-vowel scores across F0 differences [Fig. 4-1(D)]. The C_{ohc} and C_{ihc} values were reduced to fit the noise-induced hearing impaired data from cats (using the *fitaudiogram2* MATLAB[7] function

[7] The link to download this code is given below "https://www.ece.mcmaster.ca/~ibruce/zbcANmodel/zbcANmodel.htm"

implemented by Zilany et al., 2009), for a desired mean audiometric thresholds between 250 and 4000 Hz for OHI subjects (Table 4-1).

The number of CF responses from the impaired fibers in each octave segment were reduced using the parameters N_1, N_2, N_3 and N_4 of Figure 4-1(C) to include the CS effect. Compared to the ONH subjects, the effect of CS was larger for OHI subjects due to the combined effects of age and hearing-loss. The impaired AN responses collected from (Bruce et al., 2018) were passed through a modified Meddis and Hewitt (1992) F0-guided segregation algorithm with the loss of CFs to predict the concurrent vowel scores for the OHI model. For both the ONH and OHI models, it was assumed that CS has the same effect on all types of SR fibers, consistent with the modeling framework of Encina-Llamas et al. (2019), in which all types of SR fibers were required to be reduced in the AN model (Zilany et al., 2014) to simulate the effects of CS for capturing the envelope following response in humans.

4.3 Results

Figure 4-2 shows the percent correct of both vowels and the corresponding percent segregation as a function of F0 difference for the three listening models. Consistent with other YNH modeling studies (Chintanpalli and Heinz, 2013; Meddis and Hewitt, 1992; Settibhaktini and Chintanpalli, 2020), the predicted scores (solid line) increased with F0 difference and then asymptoted at 6-Hz F0 difference [Fig. 4-2(A)]. This identification score pattern is similar to the Chintanpalli et al. (2016) concurrent vowel data for YNH subjects (dashed line). The percent F0 segregation (i.e., ability to utilize F0-guided segregation prior to identification) improved with increasing F0 difference and reached the maximum at higher F0 differences [Fig. 4-2(E)]. The improvement in the identification of both vowels [Fig. 4-2(A)] can be attributed to enhancement in percent segregation with increasing F0 difference [Fig. 4-2(E)] (Chintanpalli and Heinz, 2013; Settibhaktini and Chintanpalli, 2020).

Table 4-2 shows the parameters that were used in the current study to successfully predict the concurrent vowel scores across F0 differences for the three listening models. To obtain the YNH model scores, the parameters *HSRsps*, *MSRsps* and *LSRsps* were set to 50, 4 and 0.1 spikes/sec, respectively, which resulted in a relative PTA difference = 9.4% (< 10%). Additionally, the F0-segregation parameter and *m1/m2* criterion were 85% and 2, respectively, with the CF-dependent time constant shown in Table 4-2. These parameter values were the same for the ONH and OHI models (Table 4-2) to evaluate how peripheral changes affect concurrent-vowel scores across F0 differences.

Table 4-2 Model parameters used across the three different listening models. SR parameters are given in spikes/sec.

Models			YNH	ONH	OHI
Parameters	AN Model	*HSRsps*	50		
		MSRsps	4		
		LSRsps	0.1		
		C_{ohc}	1	0.85	varied
		C_{ihc}	1	0.85	varied
	CS (Number of CFs in each octave segment, from low to high CFs)	N_1	25	4	2
		N_2	25	3	1
		N_3	25	3	1
		N_4	25	20	16
	Segregation algorithm	$\Delta\tau$	Same		
		F0-seg	85		
		m1/m2	2		

To assess whether EP reduction can solely account for the reduced concurrent-vowel identification scores, Figures 4-2(B) and 4-2(F) show the ONH model scores and percent segregation across F0 differences. The C_{ohc} and C_{ihc} values of the AN model (Bruce et al., 2018) were decremented from 1 with a step of 0.05, and it was found that when C_{ohc} = C_{ihc} = 0.85 (Table 4-2), the AN model's thresholds matched (relative PTA difference = 9%) the mean audiometric thresholds for ONH subjects. Thus, EP reduction was represented using the model parameters C_{ohc} and C_{ihc} but controlled by the mean audiometric thresholds of the ONH subjects. The model scores did not match with Chintanpalli et al. (2016) concurrent vowel data for ONH subjects [solid vs. dashed lines Fig. 4-2(B)], but the scores were closer to YNH subjects instead [Fig. 4-2(A), dashed line]. Hence, EP reduction based on audiometric threshold-matching did not solely address the reduced concurrent vowel scores for ONH subjects. This finding suggests that the inclusion of CS might be required for better predictions for the ONH model.

Figures 4-2(C) and 4-2(G) show the ONH model percent scores for both the vowels and its corresponding percent segregation as a function of F0 difference with the inclusion of the EP reduction (C_{ohc} and C_{ihc} = 0.85) and CS. The 100 CFs that were used for the YNH model [Fig. 4-1(A)] were divided into four different segments between 250 to 4000 Hz in octave steps [i.e., first segment: (250 to 500 Hz), second segment: (500 to 1000 Hz), third segment: (1000 to 2000 Hz) and fourth segment: (2000 to 4000 Hz)]. For the ONH model, the numbers of CFs across four segments were reduced systematically using the parameters N_1, N_2, N_3 and N_4 [Fig. 4-1(B)] to simulate the effect of CS. It was found that when the number of CFs in each segment was 4, 3, 3 and 20, respectively, the ONH model scores matched successfully with the concurrent-vowel data [Fig. 4-2(C)]. As F1 and F2 of the five single vowels (except F2 of /i/) were within the first three octave segments (Table 2-1), removing the CFs in these segments resulted in reduced concurrent vowels. This finding is consistent with Peterson and Barney (1952), suggesting that F1 and F2 are important for vowel identification. Thus, with an inclusion of CS (i.e., with total of 30 CFs), identification scores of both vowels across F0 differences [Fig. 4-2(C), solid line] were similar to Chintanpalli et al. (2016) concurrent vowel data for ONH subjects [Fig. 4-2(C), dashed line].

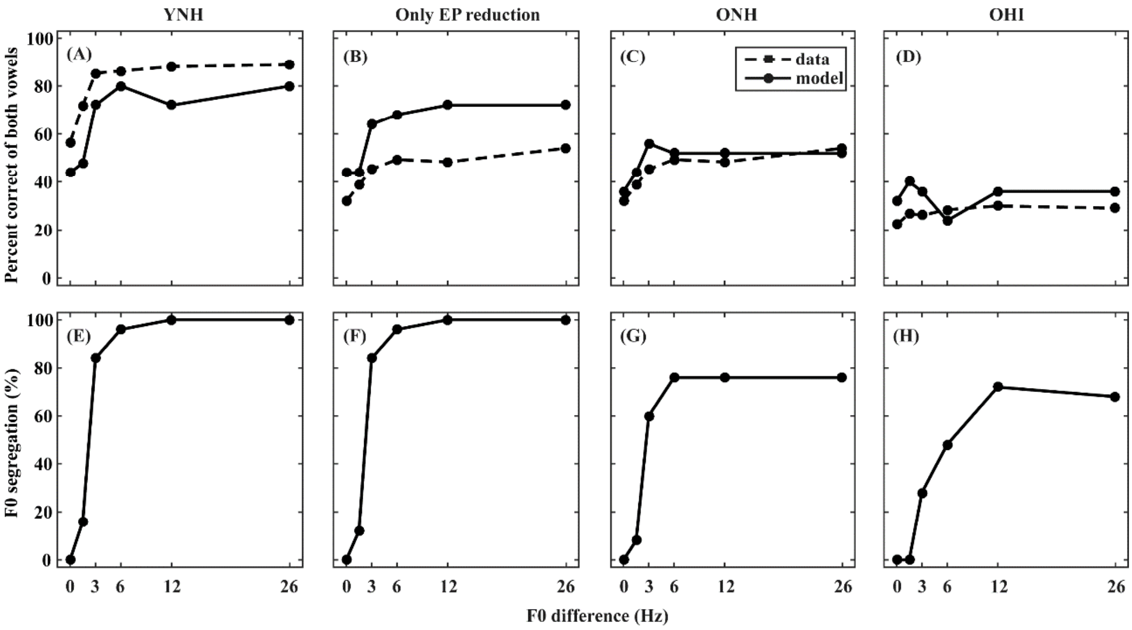

Figure 4-2 Predicted effects of F0 difference on percent concurrent vowel identification (top panels, solid lines) and percent segregation (bottom panels). Percent F0 segregation is computed as the proportion of vowel pairs (out of 25) in which the ACFs were segregated into two different sets. YNH model [panels A and E], ONH model with only EP reduction [panels B and F], ONH model [panels C and G], and OHI model [panels D and H]. The same F0-guided segregation parameters are used across models. To simulate CS, only 30 CFs (i.e., 4, 3, 3 and 20 across the four segments) and 20 CFs (i.e., 2, 1, 1 and 16 across the four segments) out of 100 are used for ONH and OHI models (third and fourth columns), respectively. For visualization purposes, the concurrent vowel data from Chintanpalli et al. (2016) are shown in the top panels using the dashed lines. The percent identification scores are shown here rather than the rationalized arcsine transformed scores, modified from Chintanpalli et al. (2016).

Figures 4-2(D) and 4-2(H) show the OHI model percent scores for both vowels and its corresponding percent segregation as a function of F0 difference due to the hearing-loss and CS. The estimated AN thresholds at audiometric frequencies, obtained after fitting, were matched (relative PTA difference = 1.45%) with the mean thresholds for OHI subjects. The C_{ohc} and C_{ihc} values were assumed to produce 2/3 OHC loss (in dB) and 1/3 IHC loss at each CF, based on hair cell damage being the dominant effect (Wu et al., 2020a, 2020b). This 2/3-OHC-loss configuration is also consistent with estimates from human listeners with mild-to-moderate SNHL (Plack et al., 2004), and NIHL AN-fiber data (Bruce et al., 2003; Miller et al., 1997). Although the estimation of C_{ohc} and C_{ihc} is based on NIHL data (and thus may not appear to include EP reduction), EP reduction can in fact be considered as being included because the NIHL values of C_{ohc} and C_{ihc} were significantly less than 0.85 used in the ONH model to capture EP reduction. This approach is consistent with recent studies suggesting the effect of reduced EP (metabolic loss) is minimal in the presence of sensory loss due to mechanical hair-cell damage (Wu et al., 2020a, 2020b) and with the idea that both hair-cell damage and reduced EP affect both OHC and IHC function, in different proportions.

The impaired AN responses based on these C_{ohc} and C_{ihc} values were obtained. Due to combined effects of age and hearing loss, a greater number of CFs were removed in the OHI model than the ONH model. Out of 30 CFs used in the ONH model, a random selection of 20 CFs resulted in reduced concurrent vowel scores. For the OHI model, the numbers of CFs across four segments were reduced systematically using the parameters N_1, N_2, N_3 and N_4 [Fig. 4-1(C)] to simulate the effect of CS. When the number of CFs across four segments were 2,1,1 and 16, respectively, the OHI model scores [Fig. 4-2(D), solid line] were matched successfully with the OHI concurrent vowel data [Fig. 4-2(D), dashed line] of Chintanpalli et al. (2016). Hence, the model fitted the lower scores for

OHI subjects by reducing the functionalities of OHCs and IHCs, and the total number of CFs (i.e., to 20) in the F0-guided segregation algorithm. This finding further highlighted that when the AN fibers closer to F1 and F2 of each vowels are removed , then the OHI model scores are reduced, suggesting that F1 and F2 are important for vowel identification (Peterson and Barney, 1952).

The percent F0 segregation was lower for the ONH model [Fig. 4-2(G)] and lowest for the OHI model [Fig. 4-2(H)] as a function of non-zero F0 difference, when compared with the YNH model [Fig. 4-2(E)]. This finding indicates that the lower scores for ONH and OHI models could be attributed to reduction in percent segregation with increasing F0 difference. However, for 0-Hz F0 difference, the scores for both the models were lower than the YNH model, suggesting that the formant difference cues are also reduced for identification. The predicted F0 benefit was highest for the YNH model and lowest for the OHI model with a moderate F0 benefit for the ONH model. Table 4-3 shows the F0 benefit comparison between concurrent vowel data (Chintanpalli et al., 2016) and the three listening models. These predicted F0 benefits were similar to Chintanpalli et al. (2016) concurrent vowel data.

Table 4-3 F0 benefit (in percent) comparison between concurrent vowel data (Chintanpalli et al., 2016) and three listening models.

Listening group	Percent F0 benefit	
	Data	Model
YNH	32	36
ONH	22	16
OHI	6.8	4

Figure 4-3 shows the model percent correct identification scores of one vowel of the pair as a function of F0 difference. The scores were 100% across

F0 differences, regardless of the listening model used. These scores were matched qualitatively with Chintanpalli et al. (2016) concurrent vowel data [compare Fig. 4-3(A) with Fig. 4-3(B)]. Consistent with YNH subjects (Chintanpalli and Heinz, 2013; Settibhaktini and Chintanpalli, 2018, 2020)(, these findings suggest that formant difference cues may be sufficient for one-vowel correct, while the F0 difference is vital for identifying the second vowel of the pair, even for ONH (consistent with Snyder and Alain, 2005) and OHI subjects.

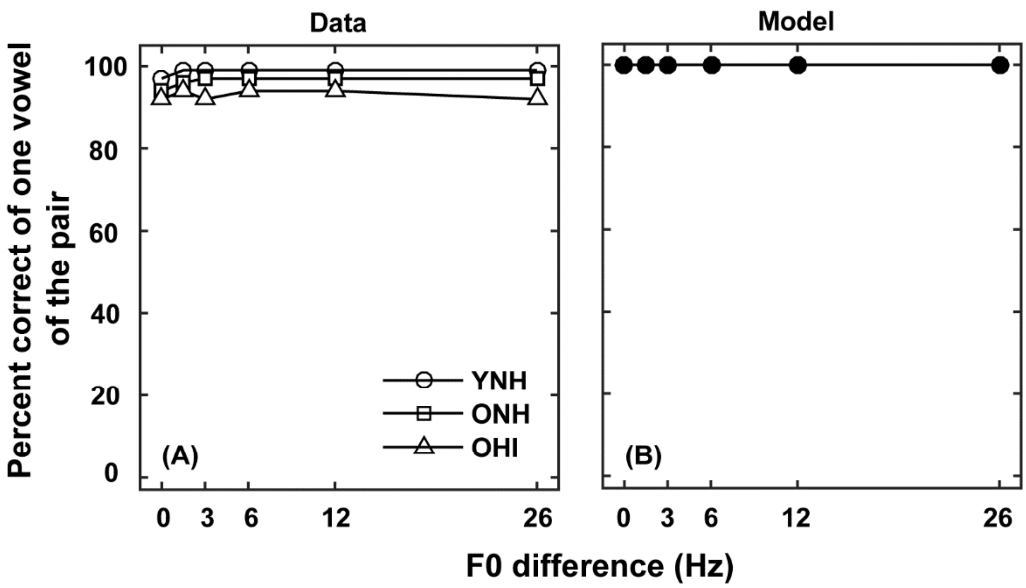

Figure 4-3 Effect of F0 difference on one-vowel-correct identification of the pair. (A) Concurrent vowel data from Chintanpalli *et al.* (2016). Note that the percent correct scores are shown instead of the rationalized arcsine transformed scores. (B) Current model predictions (all three model scores are consistently at 100%; only YNH shown).

To understand the availabilities of F0 difference and formant difference cues for identification across three listening models, the vowel pair /i, æ/ was analyzed for illustration purposes. Figure 4-4 shows the model responses to /i/ (F0 = 100 Hz), /æ/ (F0 = 106 Hz) presented to the YNH model. Figure 4-4(A) shows the individual ACF channels across 100 CFs that are logarithmically spaced between 250 and 4000 Hz. The estimated dominant F0 was correctly identified as 106 Hz [Fig. 4-4(D)]. The model does F0-guided segregation, as

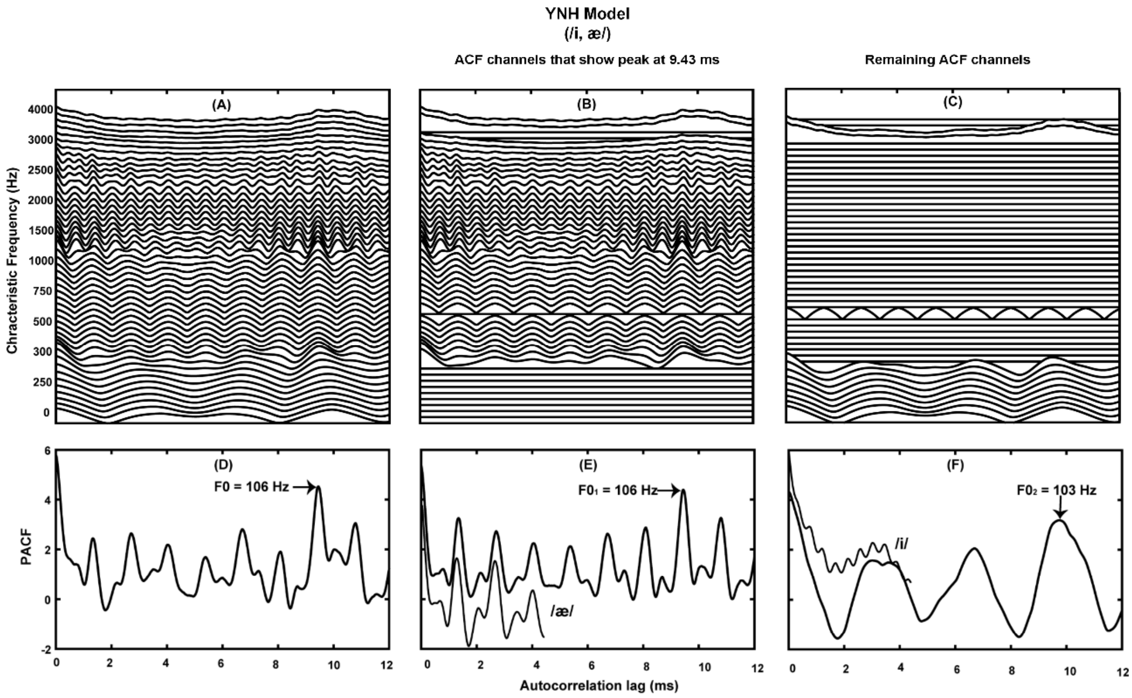

Figure 4-4 Model responses for /i/ (F0 = 100 Hz), /æ/ (F0 = 106 Hz) presented to the YNH model. The first column corresponds to the individual ACF channels from 100 different AN fibers. These channels are added together to obtain the pooled ACF [panel (D)]. The estimated dominant (F0) is 106 Hz, as indicated by an arrow [panel (D)]. The second column shows only ACF channels that have a peak at 9.43 ms [panel (B)] and the remaining channels are placed in the third column [panel (C)]. The model vowel responses are correct, as shown in panels (E) and (F). Note that the timbre regions of the templates /æ/ and /i/ (solid lines) are shown in panels (E) and (F) with an arbitrary vertical and horizontal offset for clarity. For visualization purposes, only 50% of channels are shown in the ACF plots.

only 74% of the ACF channels showed a peak at 9.43 ms (i.e., F0 = 106 Hz), which was lesser than the segregation parameter (85%). The ACF channels that had a peak at 9.43 ms were grouped together [Fig. 4-4(B)], whereas the remaining channels were grouped separately [Fig. 4-4(C)]. Two individual segregated PACFs were computed from these groups for vowel identification. The timbre region of these segregated PACFs was compared with the previously stored templates of five single vowels. The model identified /æ/ and /i/ correctly [Figs. 4-4(E) and 4-4(F)]. The individual F0 was identified correctly for /æ/ [Fig. 4-4(E)] but not for /i/ [Fig. 4-4(F)]. The model successfully identified both the vowels, as only one correct estimation of F0 is required for vowel segregation and identification (Meddis and Hewitt, 1992).

Figure 4-5 shows the model responses for the same vowel pair /i/ (F0 = 100 Hz), /æ/ (F0 = 106 Hz), but presented to the ONH model. The dominant F0 was identified correctly, as there was a peak at 9.43 ms in the pitch region of PACF [Fig. 4-5(D)]. Only 66.66% of 30 channels showed a peak at 9.43 ms and thus, F0-guided vowel segregation was allowed before vowel identification. In this case, the model identified both the vowels and individual F0s correctly [Figs. 4-5(E) and 4-5(F)]. Figures 4-4 and 4-5 suggest that the F0-guided vowel segregation was beneficial for this vowel pair for correct identification in both the YNH and ONH models.

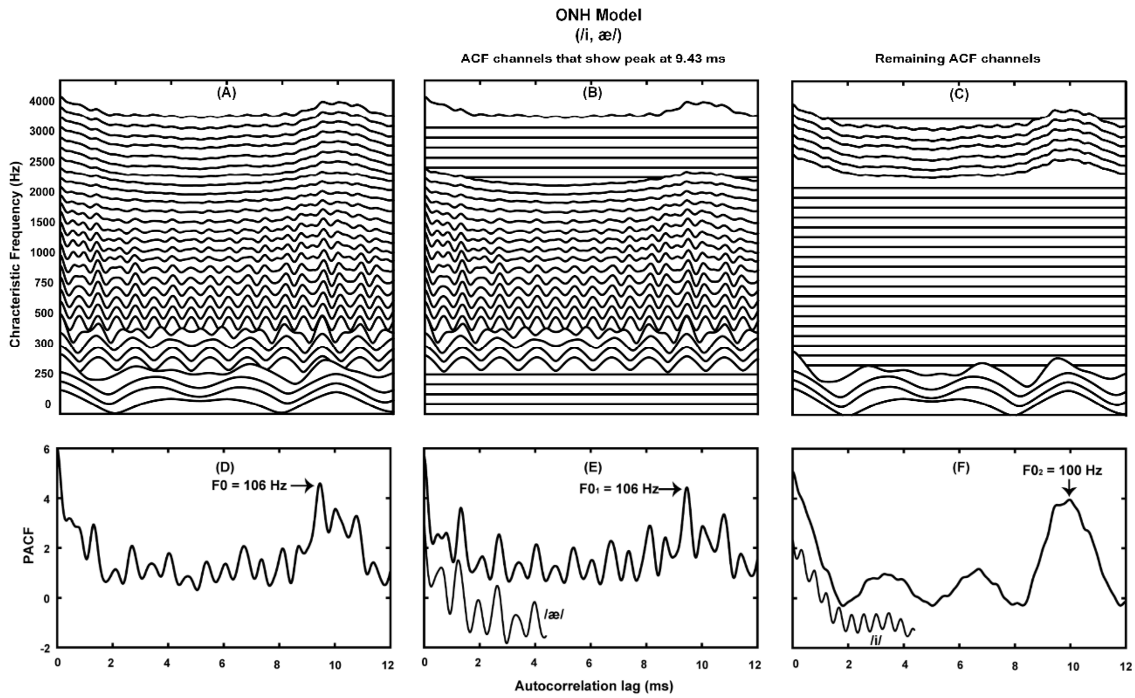

Figure 4-5 Model responses for /i/(F0 = 100 Hz), /æ/ (F0 = 106 Hz) presented to the ONH model. The first column corresponds to the individual ACF channels of the 30 selected AN fibers due to cochlear synaptopathy. The figure caption is similar to Fig. 4-4. The estimated dominant F0 is 106 Hz, as indicated by an arrow [panel (D)]. The model vowel responses are correct, as shown in panels (E) and (F).

Figure 4-6 shows the model responses for the same vowel pair /i/ (F0 = 100 Hz), /æ/ (F0 = 106 Hz), but for the OHI model. The peak in the pitch region of the PACF occurred at 9.43 ms, which corresponds to the correct estimation of F0 = 106 Hz [Fig. 4-6(B)]. Here, the model did not segregate based on F0

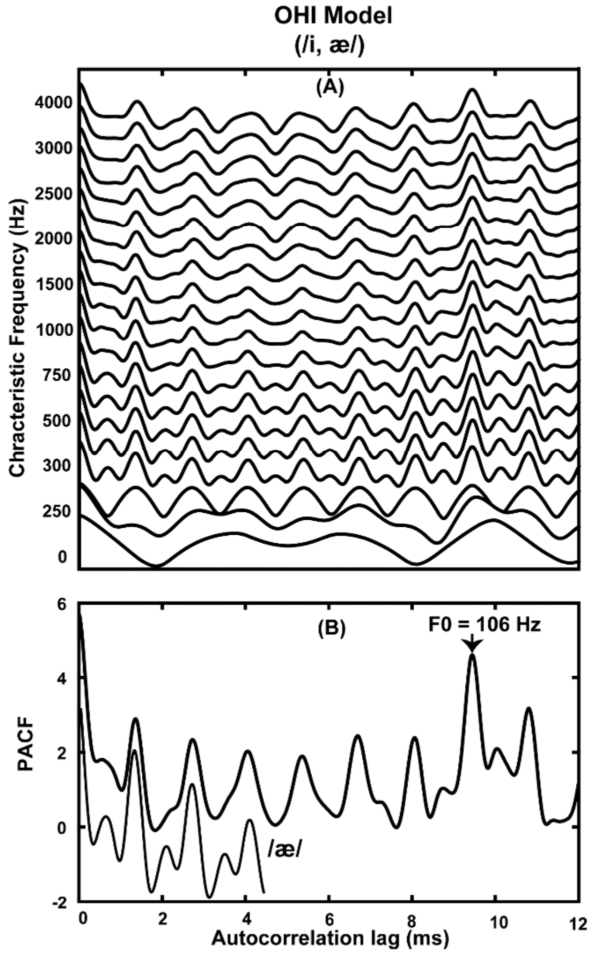

Figure 4-6 Model responses for /i/ (F0 = 100 Hz), /æ/ (F0 = 106 Hz) presented to the OHI model. The first column corresponds to 20 different ACF channels due to cochlear synaptopathy. These channels are added together to obtain the pooled ACF, as shown in the bottom panel (B). The estimated F0 is 106 Hz, as indicated by the arrow [panel (B)]. The model predicts an incorrect vowel response /æ, æ/ using no-F0 segregation condition. Note that the timbre regions of the templates /æ/ (solid line) is shown in panel (B) with an arbitrary vertical and horizontal offset for clarity.

difference, as the percent ACF channels that showed a peak at 9.43 ms was 90% (> 85% segregation parameter). The model incorrectly identified /æ, æ/ based on *m1/m2* criterion and predicted a single F0 = 106 Hz. However, the model can predict the correct vowel pair only when the *m1/m2* was increased to 16.67. Due to the lack of the F0-guided segregation cue for the OHI model [Fig. 4-2(H)], vowel identification was based on the formant difference cues, which resulted in incorrect identification.

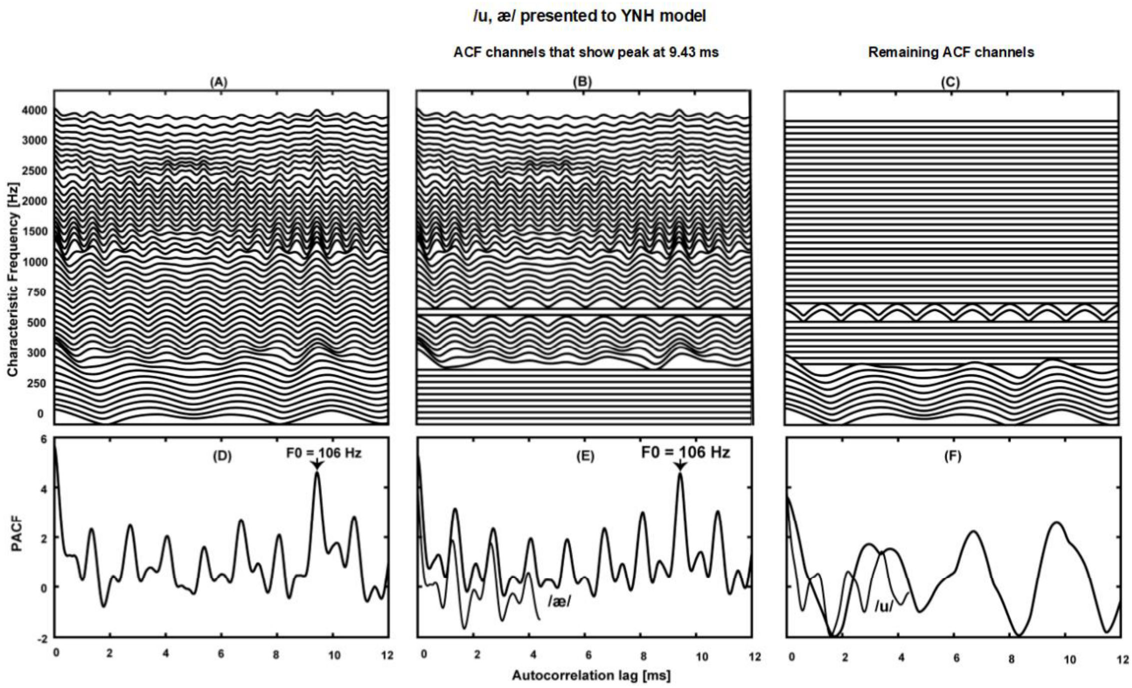

Figure 4-7 Model responses for /u/ (F0 = 100 Hz), /æ/ (F0 = 106 Hz) presented to the YNH model. The figure caption is similar to Fig. 4. The estimated F0 = 106 Hz. The second column shows only ACF channels that have peak at 9.43 ms [panel (B)] and the remaining channels are placed in the third column as shown in panel (C). The model vowel responses are correct, as shown in panels (E) and (F). The timbre regions of the templates /æ/ and /u/ (solid lines) are shown in panels (E) and (F) with an arbitrary vertical and horizontal offset for clarity.

To illustrate how the lack of the F0-guided segregation cue (i.e., with only formant-difference cues) for both ONH and OHI models resulted in incorrect responses, another vowel pair /u/ (F0 = 100 Hz), /æ/ (F0 = 106 Hz) was also analyzed. Figure 4-7 shows the model responses of another vowel pair /u/ (F0 = 100 Hz), /æ/ (F0 = 106 Hz) presented to the YNH model. Similar to Figure 4-4, the model identified both the vowels correctly based on the segregated PACFs. Figure 4-8 shows the model responses for same vowel pair /u/ (F0 = 100 Hz), /æ/ (F0 = 106 Hz), but for the ONH and OHI models. There was a peak at 9.43 ms for both the models that corresponded to 106 Hz [Figs. 4-8(C) and 4-8(D]. Additionally, the percent ACF channels that showed a peak at 9.43 ms was 86.66% and 90% for ONH and OHI models [Figs. 4-8(A) and 4-8(B)], respectively. Thus, the model did not segregate using the F0 difference but had to identify the vowels based on the *m1/m2* criterion. The model incorrectly identified /æ, æ/ and predicted a single F0 = 106 Hz. If the model (ONH or OHI) had to pick two different vowels, an incorrect response /i, æ/ was identified. Due to the lack of F0-guided segregation in both models [Figs. 4-2(G) and 4-2(H)], the vowel pair was incorrectly identified, as the identification was based on formant-difference cues.

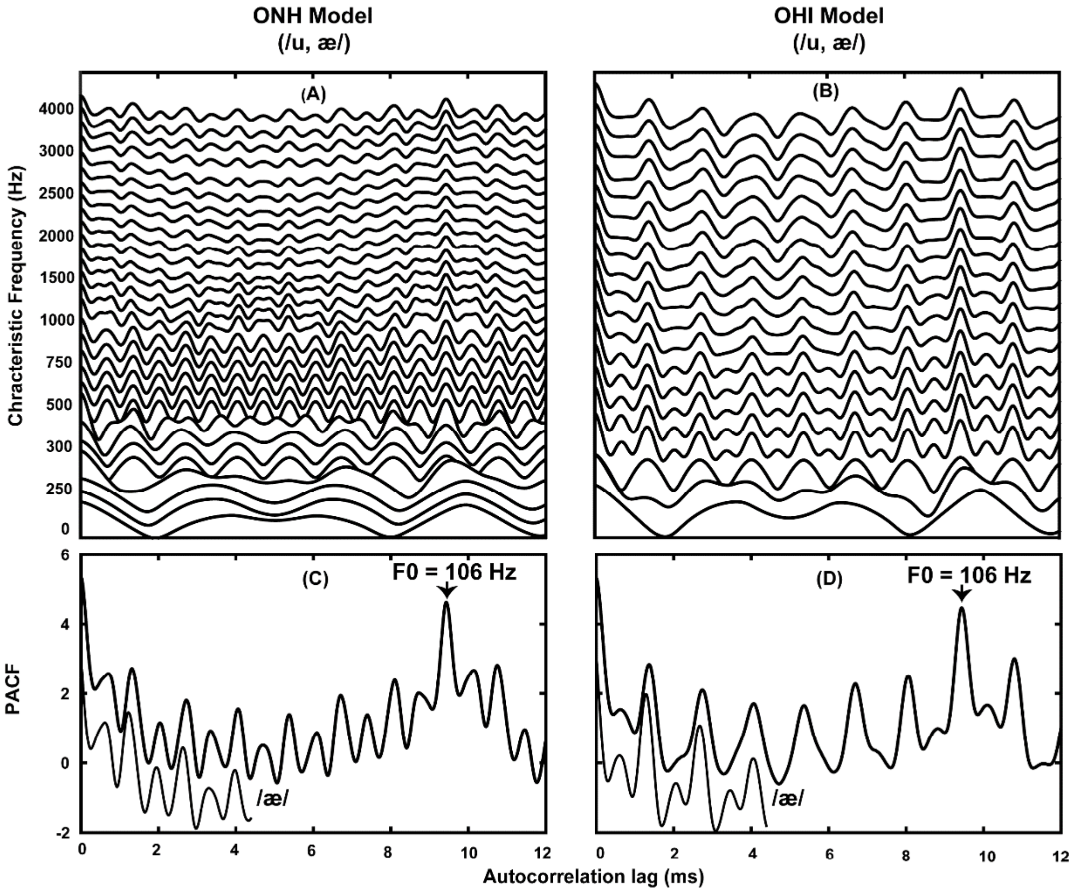

Figure 4-8 Model responses for /u/ (F0 = 100 Hz), /æ/ (F0 = 106 Hz) presented to the ONH model (first column) and OHI model (second column). Top panels (A-B) shows the individual ACF channels computed from 30 and 20 different AN fibers, respectively. These channels are added together to obtain the pooled ACF, shown in bottom panels (C-D). The estimated F0 is 106 Hz, as indicated by the arrow in the bottom panels. The model predicts an incorrect answer /æ, æ/ under the no-F0 segregation condition. The timbre regions of the /æ/ (solid line) are shown in the bottom panels.

4.4 Discussion

4.4.1 Effect of F0 difference cue on concurrent vowel scores across three listening models

The model predictions for the three different listening models [Fig. 4-2(A) for YNH model, Fig. 4-2(C) for ONH model and Fig. 4-2(D) for OHI model] were successful in capturing the pattern of concurrent-vowel identification scores across F0 differences. The F0 benefits were similar between the concurrent-vowel data and the model scores for each listening model (Table 4-3). The percent segregation was reduced with increasing age, suggesting that the F0-guided segregation ability was limited, which correlated well with the ONH model's reduced identification scores [compare Figs. 4-2(A) with 4-2(C)]. This segregation was further reduced due to hearing loss, resulting in the lowest identification scores for the OHI model [compare Figs. 4-2(C) with 4-2(D)]. Our modeling predictions suggest that the limited vowel segregation based on F0 difference for ONH and OHI subjects might contribute to reduced concurrent-vowel scores. For each listening model, the one vowel correct identification was 100% and was qualitatively successful in matching with the concurrent vowel data [compare Figs. 4-3(A) with 4-3(B)], suggesting that F0 difference is essential for the identification of the second vowel in the pair.

4.4.2 Effect of CS on predicting concurrent-vowel scores for the ONH and OHI models

The CS is an important anatomical change that occurs in the periphery due to normal aging and also with hearing loss but has a negligible effect on the audiometric profile of the listeners (Liberman and Kujawa, 2017). The effect of CS on perception is yet to be fully understood. Thus, the current study attempts to evaluate this effect by predicting concurrent-vowel scores for ONH and OHI subjects. The modeling predictions confirm that concurrent-vowel scores are reduced for the ONH model by including CS [compare Fig. 4-2(B) with Fig. 4-2(C)]. Furthermore, the OHI modeling predictions [Fig. 4-2(D)] were lowest primarily due to CS (hearing impairment also contributed). The current study successfully captured the pattern of ONH and OHI identification scores across F0 differences, when the total number of CFs was reduced to 30 and 20 (out of 100), respectively.

The current study assumed a loss of all AN-fiber SR types at each CF to simulate the CS effect for the ONH and OHI models. The rationale for this approach was to use as simple a model as possible to represent the basic loss of AN fibers with CS, without being more specific than current data justify. While the CS data from rodents suggest that low-SR fibers are lost in greater proportion that high-SR, whether this occurs in humans remains unknown. In fact, model simulations that capture the effect of CS on envelope following responses in humans required loss of all SR types to account for the human data (Encina-Llamas et al., 2019). By having 25 CFs per octave in the YNH model, reducing the total number of AN fibers within each octave is a simple way to capture the main effect of having fewer AN fibers to convey information to the central processing structures.

4.4.3 Sensitivity of predicted concurrent-vowel scores to model parameters

It is assumed that OHC and IHC dysfunction are equal in response to EP reduction in the ONH model. The C_{ohc} and C_{ihc} values were estimated by matching the thresholds of the AN model with the mean audiometric thresholds for ONH subjects. The scores for ONH model, solely based on C_{ohc} and C_{ihc} = 0.85, had a minimal effect on the concurrent vowel scores, relative to YNH model [compare Figs. 4-2(A) vs. 4-2(B)]. We also investigated the sensitivity of ONH-model scores to setting C_{ohc} and C_{ihc} to be equal for representing EP reduction.

It was found that when C_{ohc} = 0.8 and C_{ihc} = 0.85, there was still a good fit (relative PTA difference < 10%) between the thresholds of AN and mean audiometric thresholds of ONH subjects. However, with this change the concurrent-vowel scores were the same as shown in Figure 4-2(B) (result not shown). These findings suggest that the ONH model scores are not particularly sensitive to exactly equal contributions of OHC and IHC for reducing the EP.

The values for the *HSRsps*, *MSRsps* and *LSRsps* fibers of the AN model were 50, 4 and 0.1 spikes/sec, respectively. These values were selected, such that the thresholds of the AN model matched well with the mean audiometric thresholds for YNH subjects (Chintanpalli et al., 2016) with relative PTA difference (< 10%). Hence, there is also a possibility that an alternative set of SR values (spikes/sec) could be obtained to match the mean thresholds of YNH subjects. For that alternative set, the absolute values of parameters of the F0-guided segregation algorithm may have to be adjusted for the YNH model to fit the effect of F0 difference on concurrent-vowel identification [similar to Fig. 4-2(A)]. Regardless of these changes, as long as the ONH and OHI models were developed based on the approaches proposed in the current study (Fig. 4-1), the models' scores will be able to provide a good fit to concurrent-vowel data.

It was assumed that OHI subjects had NIHL, with 2/3 OHC dysfunction (in dB) and 1/3 IHC dysfunction at each CF, values that are consistent with SNHL data in both animal and human (Bruce et al., 2003; Plack et al., 2004). It is possible that the OHI subjects could have a different configuration of OHC and IHC losses at each CF, however, the current modeling represents a simple approach in assuming the same proportion of OHC to IHC loss at all CFs given that it is currently difficult to estimate this proportion in individual subjects. Furthermore, our approach assumes that these values include both the EP-reduction metabolic effects (equal reduction in C_{ohc} and C_{ihc}) along with the sensory effects (larger reduction in C_{ohc} and C_{ihc}, consistent with suggestions from recent temporal bone studies in humans, Wu et al., 2020a, 2020b). As techniques become available to define the specific effects of EP-reduction and sensory dysfunction on OHC and IHC in individual subjects, impaired AN responses (obtained using the newly estimated C_{ohc} and C_{ihc} values across CFs) could be used to refine the OHI model to predict the concurrent-vowel scores.

4.4.4 Possible physiological mechanisms underlying reduced concurrent vowel scores due to increased age and hearing loss

In the YNH model, the concurrent vowel score was increased with an improvement in percent segregation as a function of F0 difference [Figs. 4-2(A) and 4-2(E)]. This suggests that enhanced segregation ability with F0 difference could be associated with an improvement in the strength of phase locking of AN fibers to at least one F0 of the vowel pair. The better segregation might also improve the phase locking to formants of the concurrent vowel, thereby aiding in identification. In contrast, the lower identification scores for ONH and OHI models appear to be due to the degraded phase locking of AN fibers to F0s and formants of the vowel pairs, with a larger reduction for the OHI model.

4.4.5 Conclusion

To understand the effects of age and hearing loss on the ability to utilize the F0-difference cue, the current modeling study predicted concurrent vowel scores across F0 differences for three different listening groups (YNH, ONH and

OHI). The YNH model was developed by cascading a physiologically realistic AN model (Bruce et al., 2018) with a modified Meddis and Hewitt (1992) F0-guided segregation algorithm [Fig. 4-1(A)]. The scores for the YNH model were successful in capturing the pattern of identification scores of concurrent-vowel data [Fig. 4-2(A)]. To assess age effects, EP reduction and CS were included [Fig. 4-1(B)]. The identification scores were reduced across F0 differences for the ONH model and were successful in capturing the pattern of identification scores of concurrent-vowel data [Fig. 4-2(C)]. For the OHI model, hair-cell damage, EP loss and CS (with a larger effect than for the ONH model) were incorporated [Fig. 4-1(C)], which resulted in the lowest predicted concurrent-vowel scores and captured the pattern of identification scores of concurrent-vowel data [Fig. 4-2(D)]. The reduced vowel segregation in both the ONH and OHI models [Figs. 4-2(G) and 4-2(H)] suggest a limited use of the F0-difference segregation cue with increased age and hearing loss. Additionally, all three models successfully captured the pattern of one-vowel correct and the F0 benefit observed in the concurrent-vowel data (Table 4-3). These model predictions support our hypotheses and suggest that peripheral changes due to increased age and hearing loss could contribute to reduced concurrent-vowel scores for ONH and OHI subjects.