2 Modeling the level-dependent changes of concurrent vowel scores[1,2]

2.1 Introduction

Younger adults with normal hearing (YNH) have a remarkable ability to segregate two or more simultaneous speech sounds, when presented at varying sound levels. They utilize many cues to identify the speech during this multi-talker scenario. Some of these cues are the differences in onset and offset timing of the speech sounds, differences in speech characteristics (e.g., formant differences), and differences in F0 between the speakers (e.g., Bregman, 1990; Brokx and Nooteboom, 1982; Cherry, 1953; Zwicker, 1984). Among these, F0 difference is widely studied as an important cue for segregating multiple speech signals. Concurrent vowel identification is often studied to understand how F0 difference helps to identify two simultaneously presented vowels. In this experimental paradigm, two vowels with equal duration and level are presented simultaneously to one ear of a human listener. The task of the listener is to identify both vowels, and performance is measured as a function of F0 difference between the two vowels. For YNH listeners, identification scores for both vowels improve as F0 difference increases between the two vowels, and then asymptotes usually at ~3-Hz F0 difference or higher (Assmann and Summerfield, 1990; Summerfield and Assmann, 1991; Culling and Darwin, 1993; Arehart et al., 1997, 2005; Summers and Leek, 1998; Vongpaisal and Pichora-Fuller, 2007;

[1] This work was published in the Journal of the Acoustical Society of America, vol. 143(1), 440 – 449, 2018. doi:10.1121/1.5021330

[2] The preliminary portion of this work was published in the 40th meeting of Association of Research in Otolaryngology in Baltimore, USA, 2017.

Chintanpalli and Heinz, 2013; Chintanpalli et al., 2016). However, these studies have been typically conducted at a single vowel level.

To understand how the ability to utilize the F0 difference cue varies across sound levels, Chintanpalli et al. (2014) collected concurrent-vowel data for 0 and 26-Hz F0 difference conditions (i.e., same and different F0, respectively) in YNH listeners (generally between 20 and 26 years). Figure 2-1 shows the percent correct identification scores as a function of vowel level for same and different F0 conditions (Chintanpalli et al., 2014). Their subjects showed an improvement in percent identification score as vowel level was increased from low-to-medium (25 - 50 dB SPL) and then a decline was observed at higher levels (65 – 85 dB SPL). The F0 benefit is defined as the difference in percent correct identification scores between the different and same F0 conditions, and is commonly used in the concurrent-vowel literature (Arehart et al., 2005; Assmann and Summerfield, 1990; Chintanpalli et al., 2014, 2016; Summers and Leek, 1998; Vongpaisal and Pichora-Fuller, 2007). The mean percent of F0 benefit increased from 25 to 50 dB SPL and then remained fairly constant from 50 to 85 dB SPL.

In order to understand the neural mechanisms underlying these level-dependent identification scores, Chintanpalli et al. (2014) further performed computational modeling by quantifying the phase-locking of AN fibers to vowel formants and F0s using the ALSR (section 1.1.3.2) (Young and Sachs, 1979) and template contrast (Larsen et al., 2008), respectively. These two neural coding schemes were computed from the responses of a well-established AN model (Zilany et al., 2009). It was inferred that the F0-difference cue was degraded at 25 dB SPL because this level had the lowest F0 benefit, which could

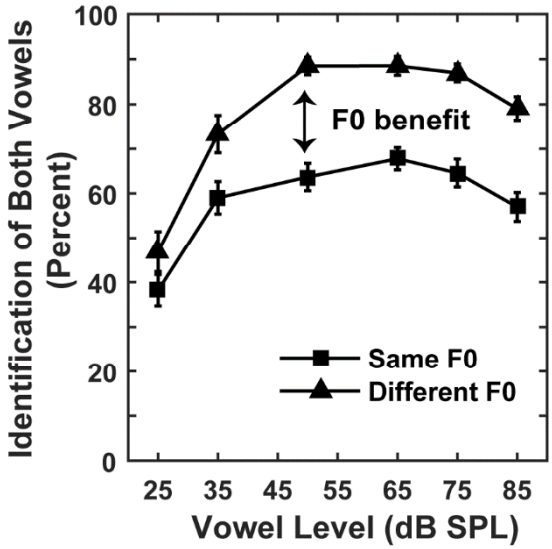

Figure 2-1 Percent identification scores of both the vowels for same F0 (squares) and different F0 (triangles) as a function of the vowel level. Note that the figure shown is the percent identification scores rather than the rationalized arcsine transformed scores, modified from Chintanpalli *et al*., (2014). Error bar denotes ± 1 Standard Error Mean (SEM). The F0 benefit is indicated by an arrow.

be attributed to poor phase- locking (or rate place coding) of AN fibers to F0s. Furthermore, the reduced scores at higher levels were attributed to poor phase-locking to formants (particularly the second formant), resulting from broader auditory filters at higher SPLs. It was assumed that the phase-locking of AN fibers to formants and to F0s contribute separately to vowel identification. However, previous modeling studies (Chintanpalli and Heinz, 2013; Meddis and Hewitt, 1992) suggest that listeners first try to segregate the vowel pair (based on the available cues), and then identify them individually. This suggests that vowel segregation precedes identification, and thus one needs to incorporate the interactions between formant and F0-difference cues for identification. However, the separate analyses of phase-locking to formants and F0s limits the direct conclusions that can be made from the modeling work of Chintanpalli et al. (2014) in terms of the level-dependent changes in identification scores of both vowels. Additionally, their conclusions were based on two concurrent-vowel pairs (/ɑ, æ/ and /i, æ/) and did not compute overall identification scores . Hence, this

limited modeling approach only allowed inferences to be made that the level-dependent changes in neural coding schemes (i.e., ALSR and template contrast) may have sufficient information to predict the pattern of concurrent-vowel scores for YNH subjects. These conclusions would have been stronger had this study computed the overall identification scores.

The aim of the first objective in this dissertation is to significantly extend the modeling work done in Chintanpalli *et al*. (2014) to test explicitly their conclusions, which were based on neural coding schemes. Here, we compute the identification scores using an improved version of the AN model (Zilany et al., 2014) and a well-established F0-guided segregation algorithm (Meddis and Hewitt, 1992). This segregation algorithm takes into account the interactions between formant and F0 difference cues and computes the identification scores for concurrent vowels. A similar type of modeling framework has also been used in previous studies to successfully capture (at least qualitatively) the effect of F0 difference on vowel identification for a given level (Chintanpalli and Heinz, 2013; Meddis and Hewitt, 1992). The aim of the present study was to examine whether this modeling framework can successfully predict the level-dependent changes in identification scores of both vowels across various sound levels and for same- and different-F0 conditions.

2.2 Methods

2.2.1 Stimuli

The stimulus generation for the current study was similar to those reported in Chintanpalli et al. (2014). A set of five different vowels (/i/, /u/, /ɑ/, /æ/ and /ɜ˞/) were generated using a MATLAB implementation of a cascade formant

synthesizer (Klatt, 1980). The duration of each vowel was 400 ms, including 15-ms raised-cosine rise and fall ramps. The formant frequencies and the bandwidth of each vowel are shown in Table 2-1, which is similar to the earlier studies on concurrent vowel identification (e.g., Assmann and Summerfield, 1994; Summers and Leek, 1998; Chintanpalli and Heinz, 2013; Chintanpalli et al., 2014; Chintanpalli et al., 2016).

A concurrent vowel was obtained by adding any two individual vowels. To form a vowel pair for different 26-Hz F0 difference condition (different F0), one vowel had a constant F0 = 100 Hz, whereas another vowel had F0 = 126 Hz. These five vowels were arranged in different combinations to obtain 25 concurrent vowels. To maintain equivalent numbers of vowel pairs, 0-Hz F0 difference condition also included 25 vowel pairs (five identical-vowel pairs and ten different-vowel pairs, but the latter was presented twice, where F0 was 100 Hz for both vowels in the pair). A total of 50 concurrent vowels (25 vowel pairs x two F0 conditions) were used at each level. The individual vowel levels are 25, 35, 50, 65, 75, and 85 dB SPL, respectively. At each level, the vowel pairs either had same (100 Hz) or different F0s (100 and 126 Hz). Overall, 300 vowel pairs were generated (25 vowel pairs x 2 F0 conditions x 6 levels).

Table 2-1 Formants in Hz for five different vowels. Values in parenthesis of the first column correspond to bandwidth around each formant (in Hz).

Vowel	/i/	/	/u/	/æ/	/ɜ˞/
F1 (90)	250	750	250	750	450
F2 (110)	2250	1050	850	1450	1150
F3 (170)	3050	2950	2250	2450	1250
F4 (250)	3350	3350	3350	3350	3350
F5 (300)	3850	3850	3850	3850	3850

2.2.2 Computational Modeling: Predicting identification scores across sound level and F0-difference conditions

To understand the level-dependent changes underlying the peripheral processing associated with the identification scores of both vowels (Fig. 2-1), the first objective utilizes the computational model (Fig. 2-2) by cascading the AN model (Zilany et al., 2014) with the modified version of F0-guided segregation algorithm (Meddis and Hewitt, 1992). The Meddis and Hewitt (1992) F0-guided segregation algorithm is the only algorithm that has successfully captured (at least qualitatively) the effect of F0 difference on concurrent vowel identification.

2.2.2.1 Auditory-nerve model

The AN model developed by Zilany et al. (2014) was used to predict the AN responses to concurrent vowels. It is an extension of several previous versions of the model, which have been tested extensively against neurophysiological responses from cats to pure tones, two-tone complexes, broadband noise and vowels (Carney, 1993; Zhang et al., 2001; Heinz et al., 2001; Bruce et al., 2003; Tan and Carney, 2003; Zilany and Bruce, 2006, 2007, Zilany et al., 2009). Relevant to the current objective, this model captures (1) the level-dependent changes in cochlear nonlinearities (e.g., compression, suppression, broadened tuning and best-frequency shifts with increase in sound levels, Ruggero et al., 1997) and (2) the level-dependent changes in phase locking ability of AN fibers to vowel formants (Chintanpalli et al., 2014; Miller et al., 1997; Zilany and Bruce, 2006, 2007) and F0 coding (Chintanpalli et al., 2014).

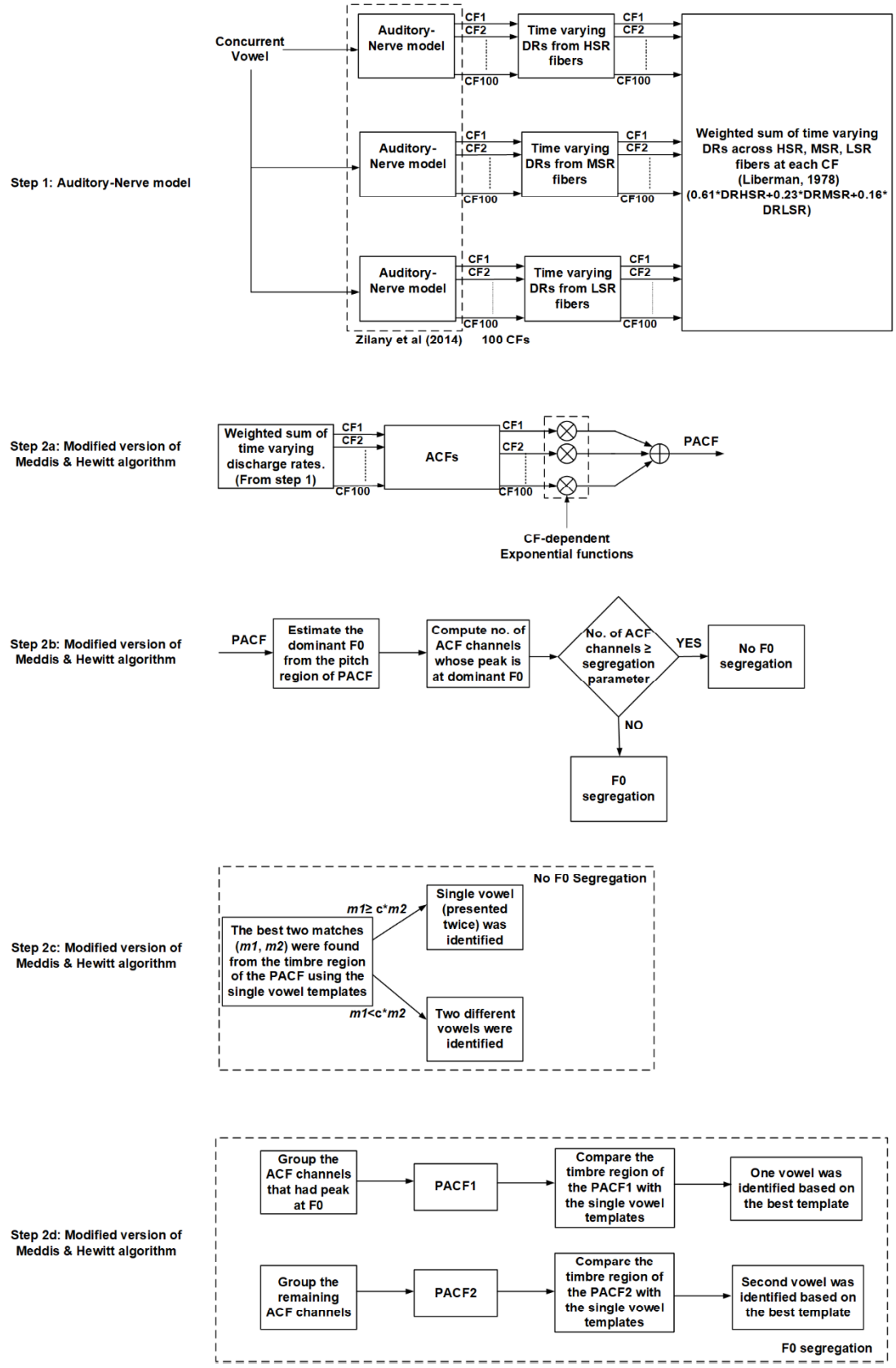

Figure 2-2 The Block diagram illustrating the steps involved in computational modeling for predicting the level-dependent changes in concurrent vowel scores for same and different F0 conditions.

The input to the AN model was a concurrent vowel and the output was a time varying DR of a single AN fiber for a particular CF. The AN responses were computed from 100 different logarithmically spaced CFs, ranging from 125 Hz – 4000 Hz, for each of the three fiber types (step 1 of Fig. 2-2). The overall DR at each CF was the weighted sum of the DRs as per the distribution of the spontaneous rates (HSR = 0.61, MSR = 0.23 and LSR =0.16; Liberman, 1978; step 1 of Fig. 2-2), which is relevant for modeling the level effects for concurrent vowels since AN-fiber threshold is inversely related to SR.

2.2.2.2 F0-guided segregation algorithm for concurrent-vowel identification

To understand the changes in identification score of both vowels across F0 difference (Fig. 1-1), there have been many attempts to develop physiologically inspired computational models that could validate the effect of F0 difference on the identification scores of both the vowels (Assmann and Summerfield, 1990; Chintanpalli and Heinz, 2013; Meddis and Hewitt, 1992). The Meddis and Hewitt (1992) model was the first to successfully capture the effect of F0 difference on concurrent-vowel identification. They showed that the identification scores of both vowels improved with increasing F0 difference and asymptoted at higher F0 differences, qualitatively. The model predictions were obtained using two stages: 1) a peripheral model, and 2) an F0-guided segregation algorithm. Subsequently, Chintanpalli and Heinz (2013) replicated this effect using the same F0-guided segregation algorithm of Meddis and Hewitt (1992) but with a more recent AN model (Zilany and Bruce, 2007). However, the parameter values of this segregation algorithm were different in each study because of the different peripheral models that were used.

In the Meddis and Hewitt (1992) F0-guided segregation algorithm, an autocorrelation function (ACF) was computed for each of the weighted sum of time-varying DR of AN fibers at each CF (step 2a of Fig. 2-2). Each ACF was multiplied by a single exponential delaying function with the time constant ($\Delta\tau$) = 10 ms. Equation 2-1 specifies the ACF computation, where $S_k(t)$ is a time-varying DR of AN fiber from the k^{th} channel, ℓ is the autocorrelation lag, t is the time at which ACF is sampled, and $\Delta\tau$ is the time constant of the exponential function. A pooled-ACF (PACF) was computed by summing across the ACF's from different fibers (step 2b of Fig. 2-2 and Equation 2-2, where n is the number of channels).

$$ACF(k,\ell) = \sum_{t=1}^{\infty} S_k(t) S_k(t+\ell) e^{-t/\Delta\tau} \qquad 2\text{-}1$$

$$PACF(\ell) = \sum_{k=1}^{n} ACF(k,\ell) \qquad 2\text{-}2$$

The dominant F0 for each vowel pair was estimated by computing the inverse delay of the largest peak in the pitch region (4.5 – 12.5 ms) of the pooled-ACF (step 2b of Fig. 2-2). It specifies for which F0 the ACFs across CFs are primarily responding. Furthermore, the computation of dominant F0 is required for deciding whether F0-guided vowel segregation can be allowed prior to the identification of individual vowels. If the number of ACF channels that showed a peak at the dominant F0 was greater than a user-defined segregation parameter, then the model decided that there was only one F0 present (i.e., no-F0 difference); otherwise, the model decided two F0s were present and proceeded with the F0-guided vowel segregation (step 2b of Fig. 2-2). If the identified

channels were greater than the user-defined segregation parameter, then the algorithm decided there was one F0; otherwise two F0s were present (step 2b of Fig. 2-2). In the latter case, all the individual ACF channels that showed a peak at the dominant F0 were summed together to obtain a pooled-ACF of one vowel of the pair (PACF1). The residual ACF channels were summed together to obtain a pooled-ACF of the other vowel (PACF2). The inverse Euclidean distance metric was then used between the timbre region (0.1 – 4.5 ms, formant frequency region) (Meddis and Hewitt, 1992) of the segregated pooled-ACF (PACF1 or PACF2) and the timbre regions of previously-stored pooled-ACF templates of single vowels. The model predicted the vowel with the maximum inverse distance from each segregated PACF (step 2d of Fig. 2-2). The single vowel template was obtained by averaging the timbre regions of the PACFs across 6 F0 conditions (i.e., 100, 101.5, 103, 106, 112 and 126 Hz). If the model predicted a single F0 (i.e., unsegregated), then the timbre regions of the PACF and five single vowels were compared using the distance metric. The model predicted a single vowel (presented twice, e.g., /æ/, /æ/) if the ratio of the best (*m1*) and second-best match (*m2*) was greater than the user-defined identification parameter; otherwise, two different vowels (e.g., /æ/, /ɑ/) were predicted (step 2c of Fig. 2-2). The percent model score at each F0 difference was computed as the proportion of vowel pairs (out of 25) in which both vowels were correctly identified in each pair. The user-defined parameter values of the segregation algorithm (i.e., ACF time constant, F0-segregation criterion, *m1/m2* criterion) were varied such that the model's scores of both vowels across F0 differences were successful in capturing the pattern of concurrent-vowel data (Assmann and Summerfield 1990).

A modified version of the Meddis and Hewitt (1992) segregation algorithm was used in the current study with the following changes. Firstly, CF-dependent time constant was used for computing the ACFs across CFs (Cariani, 2004; Bernstein and Oxenham, 2005; Chintanpalli et al., 2014). This might account for the effect of peripheral filtering on pitch perception. More specifically, smaller-bandwidth peripheral filters (i.e., at lower CFs) result in longer-duration impulse responses and thus require slower time constants. Higher-bandwidth peripheral filters (i.e., at higher CFs) result in shorter-duration impulse responses and thus require faster time constants. The value of $\Delta\tau$ was varied per CF ($\Delta\tau$ = 50 ms for $100 \leq$ CF $<$ 440 Hz; $\Delta\tau$ = 36 ms for $440 \leq$ CF $<$ 880 Hz; $\Delta\tau$ = 30 ms for $880 \leq$ CF $<$ 1320 Hz; $\Delta\tau$ = 29 ms for CF $\geq$ 1320 Hz). These CF distributions were from Cariani (2004) and the $\Delta\tau$ values were varied systematically to fit the level-dependent changes across F0 differences. Secondly, to be consistent with Meddis and Hewitt (1992), a single-vowel template at each level ('optimal case') was obtained by averaging the timbre regions of the PACFs for 100 and 126 Hz. The rationale for using a separate template for each sound level was mainly derived from the level-dependent changes due to cochlear nonlinearities.

2.3 Results

In concurrent-vowel data (Fig. 2-1), the F0 benefit increased with increasing vowel level (25-50 dB SPL) and then remained fairly constant from 50 dB SPL onwards. This suggests that the ability to segregate vowels using F0 difference cue, also increases up to 50 dB SPL. Thus, this criterion was used in the current modeling to constrain the user-defined segregation parameter of the Meddis and Hewitt (1992) algorithm across sound level. More specifically, this

parameter was selected such that 1) percent segregation ability in the two-F0 condition, computed across 25 vowel pairs, remained almost constant from 50 dB SPL onwards, and 2) percent segregation is zero for same-F0 condition at each level. Table 2-2 shows the absolute values of the segregation parameter and its corresponding F0 segregation across at each level.

To further constraint the model, *m1/m2* identification parameter was fixed to 2.5 across vowel levels. Figure 2-3(A) shows the model scores for both vowels as a function of vowel level for same (squares) and different F0 (triangles) conditions. The identification score improved as vowel level increased from low- to mid-levels and declined at higher levels for both F0 conditions. These patterns of identification scores are qualitatively similar, to that of concurrent vowel data collected from YNH subjects (Fig. 2-1).

Table 2-2 The absolute values of the segregation parameter and its corresponding percent F0 segregation at each vowel level.

Level (dB SPL)	User defined segregation parameter	Percentage of F0 segregation
25	36	48
35	43	56
50	60	84
65	52	76
75	52	76
85	56	80

Figure 2-3(B) shows the model's percent vowel segregation, computed across pairs, as a function of vowel level for same (squares) and different-F0 (triangles) conditions. As expected, the percent segregation was zero for same-F0 condition across all levels [squares in Fig. 2-3(B)]. This indicates that the level-dependent changes in identification scores for same F0 could be affected by formant-difference cues, which may be influenced largely by changes in

cochlear nonlinearities. The percent segregation increased for different F0 [triangles in Fig. 2-3(B)] until 50 dB SPL, in order to match the F0 benefit of the concurrent-vowel data, at least qualitatively.

Figure 2-3(C) shows the model scores for percent correct identification of one vowel of the pair as a function of vowel level (solid line). The identification score was 100% and was independent of F0-difference and vowel-level conditions. For comparison, this figure also shows the actual percent correct identification score of one vowel as a function of vowel level, calculated from the concurrent-vowel data (dashed line). The model score was successful in capturing the identification score of one vowel, quantitatively. These findings suggest that F0 difference and vowel level are vital for identifying the both vowels of the pair [compare Figs. 2-3(A) vs. 2-3(C) or Figs 2-1 vs. 2-3(C), dashed line]. Additionally, it may suggest that the level-dependent changes in identification scores of both vowels [Fig. 2-1 for data or Fig. 2-3(A) for model response] are largely influenced by the level-dependent changes in phase-locking of AN fibers to second-vowel characteristics (i.e., either formants or F0, or may be both). Figure 2-3(D) shows the actual (dashed line) and predicted F0 benefit (solid line) at each vowel level. Despite the model's F0 benefit being lower across levels, the model could successfully capture the pattern of variation in actual F0 benefit with vowel level, qualitatively. The lowest F0 benefit at 25 dB SPL in the model [Fig. 2-3(D)] or in the data (Fig. 2-1) could be associated with the limited F0-guided segregation [Fig. 2-3(B)] and thus confirms one of the suggestions from Chintanpalli et al. (2014).

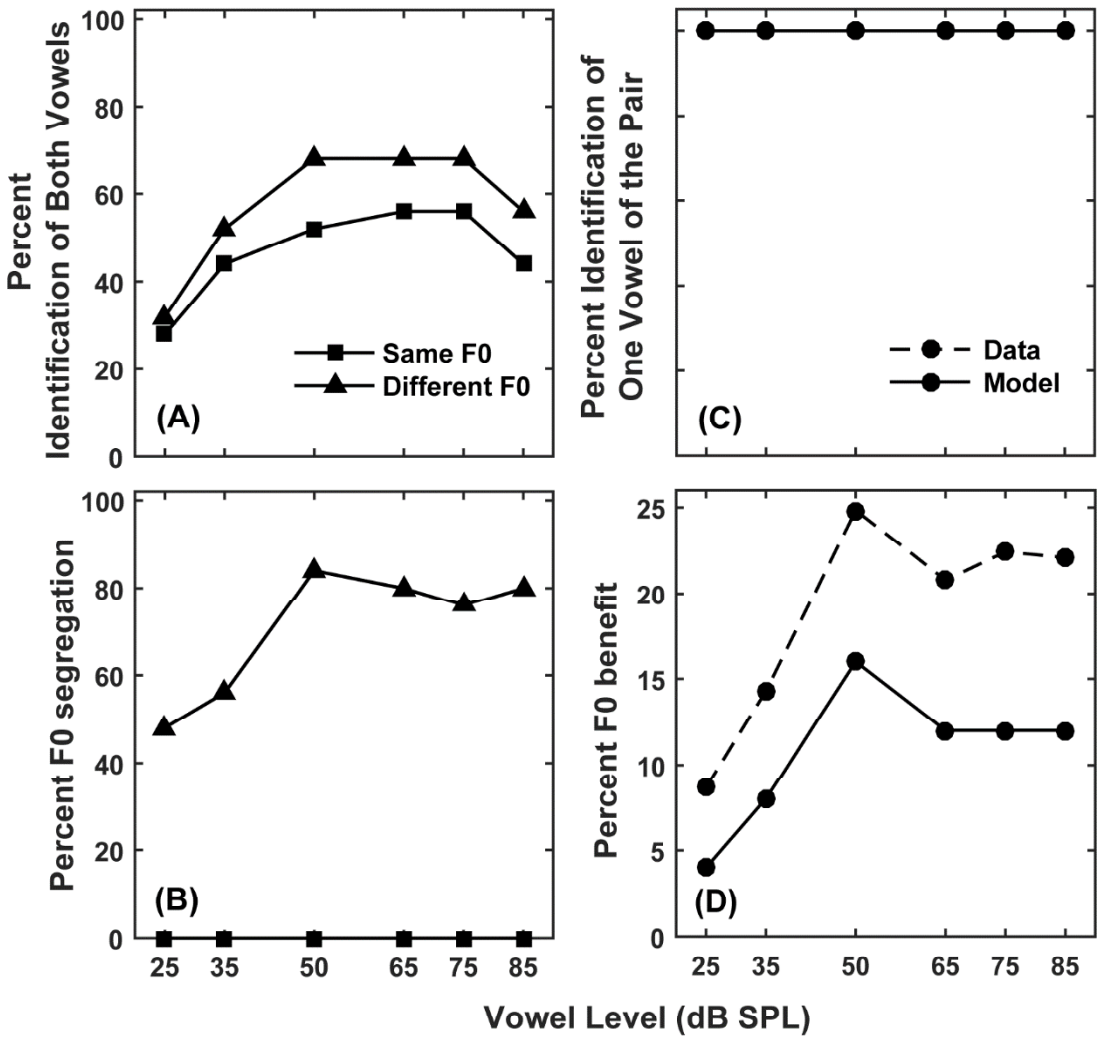

Figure 2-3 Effect of level on concurrent-vowel identification for same- and different-F0 conditions. (A) Percent identification score of both vowels, (B) Percent vowel segregation, (C) Percent identification of one vowel of the pair, and (D) F0 benefit comparison between data and model scores. Note that the legend mentioned in panel (A) also applies to panel (B) and the legend mentioned in panel (C) also applies to panel (D). The data (dashed line) corresponds to the Chintanpalli et al. (2014) behavioral data.

The model's identification score was deterministic, even with the multiple repetitions of same concurrent vowel. However, the listeners' responses include response variability that was not captured in the current modeling. To compensate this effect to some extent, the similarity scores were used to determine how well the model captures either the listeners' correct answer or confusion. This metric was also used by Chintanpalli and Heinz (2013) to evaluate their model performance in terms of listeners' responses at a single vowel level but here the current study further extends across vowel levels. As

per this metric, for a given concurrent vowel, the model was given 1 point if it predicted the correct answer, 0.5 point if it predicted the first major confusion, 0.25 point if it predicted the second confusion and 0 point if it predicted the response that was not observed in listeners' responses. Figure 2-4 shows the percent similarity score, across all pairs, as a function of vowel level for both same- (squares) and different-F0 (triangles) conditions. For the same-F0 condition, similarity scores were lower across levels. This suggests the model had the most difficulty in accounting for listeners' responses when there was no F0 guided vowel segregation. The modified version of Meddis and Hewitt (1992) segregation algorithm can account for the level-dependent changes in percent correct identification scores for same- and different-F0 conditions [Fig. 2-3(A)], but it does not fully account for concurrent-vowel identification data by listeners when specific confusions are considered.

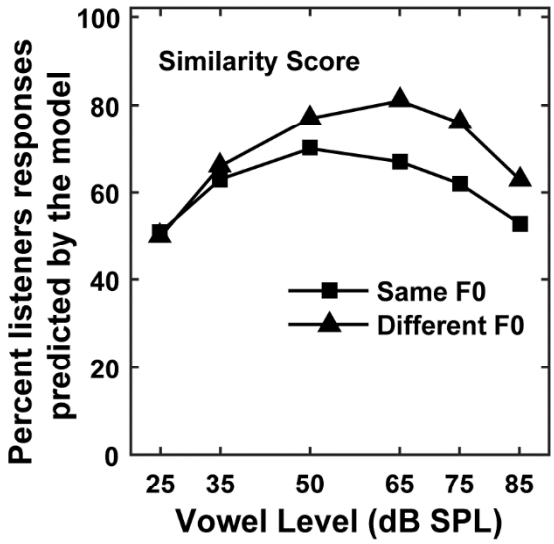

Figure 2-4 Similarity scores as a function of vowel level for same and different F0 conditions to evaluate the effectiveness of the model with respect to listeners' responses.

To understand the availabilities of formant and F0 difference cues underlying the level-dependent changes in concurrent vowel identification, a

specific vowel pair /æ, ɑ/ was analyzed for an illustration purposes. Both of these vowels have same F1 (= 750 Hz) and they differ with each other in terms of F2 and F3 (Table 2-1). Figure 2-5 shows the model response for /æ/ (F0 = 100 Hz), /ɑ/ (F0 = 100 Hz) presented at 25 dB SPL, 65 dB SPL and 85 dB SPL (individual columns). As expected, there was no F0 based segregation at each level and hence the model response was made from a single PACF using the *m1/m2* criterion (step 2 of Fig. 2-2). Figures 2-5(A-C) shows the individual ACF channels with CFs, logarithmically spaced between 125 Hz and 4000 Hz, for 25 dB SPL, 65 dB SPL and 85 dB SPL, respectively. The estimated F0 was correctly identified as 100 Hz, which was indicated by an arrow in PACF at each level [Fig. 2-5(D-F)]. For vowel identification at each level, the timbre region of the PACF was compared with the previously stored templates of five different single vowels. For 25 dB SPL, the model response was /æ, æ/, which was the second confusion (identified as 12%) in listeners' responses but the correct answer was identified as 44%. The same confusion was predicted by the model for 65 dB SPL; however, it was reduced to 7% and the correct identification score was increased to 56%. The model response was correct for 85 dB SPL but it was identified as only 28% in listeners' responses. The first confusion in this case was /æ/, /æ/ (identified as 29%) in the concurrent-vowel data and the model can successfully predict this confusion as well if the model was forced to pick only the identical vowels by increasing the *m1/m2* value. With *m1/m2* being fixed across levels, the model responses were successful in either predicting the confusion or correct answer across levels. However, only for 65 dB SPL, if *m1/m2* was changed to 2.505, then the model response was correct. If the model response had to be correct only for 25 dB SPL, then *m1/m2* should be very high

(>= 5.9). All these findings suggest that the model responses may be influenced by the level-dependent changes in cochlear non-linearities and *m1/m2* ratio of the Meddis and Hewitt (1992) F0-guided segregation algorithm.

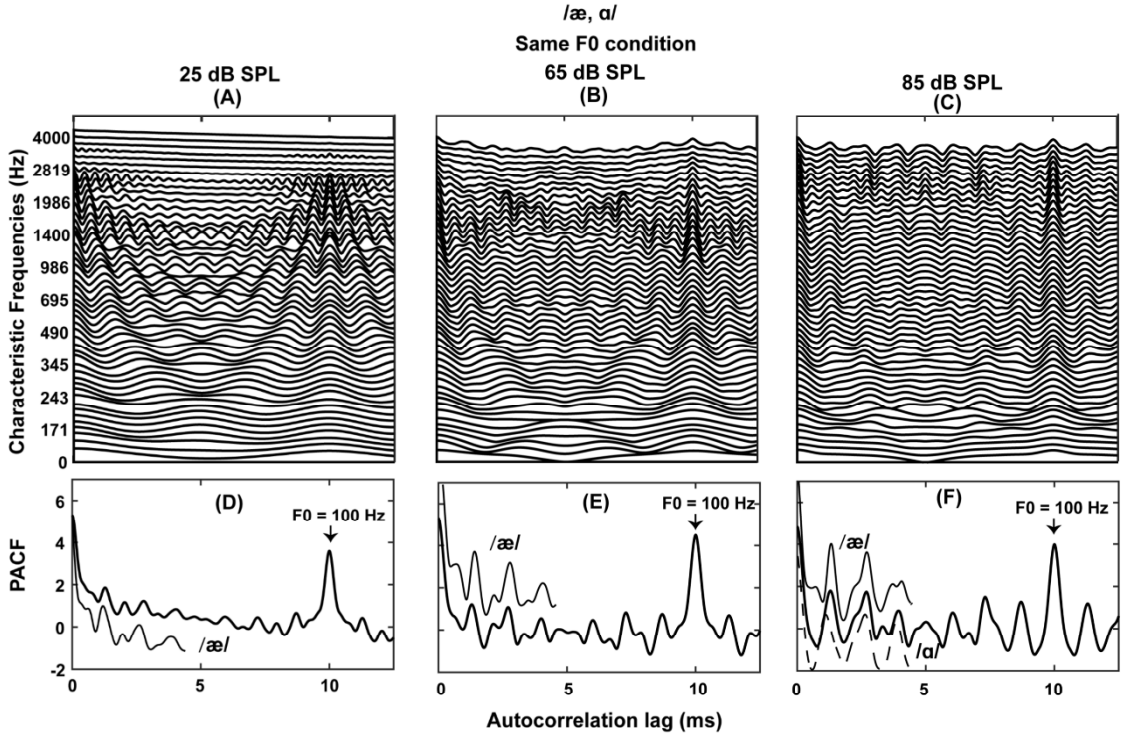

Figure 2-5 Model responses for /æ/ (F0 = 100 Hz), /ɑ/ (F0 = 100 Hz) presented at 25 dB SPL (first column), 65 dB SPL (second column) and 85 dB SPL (third column). For panels (A to C), the individual ACF channels are computed from 100 different AN fibers, ranged logarithmically between 125 Hz to 4000 Hz. These channels are added together to obtain the pooled ACF (bottom panels; D, E and F). The estimated F0 is 100 Hz, as indicated by the arrow. The model responses are shown in the bottom panels. The timbre regions of the /æ/ (solid line) are shown in the bottom panels and /ɑ/ template (dotted line) is shown only in panel F. Note that these templates are shown with an arbitrary vertical and horizontal offsets for clarity. For visualization purposes, only 50% channels are shown in the ACF plots.

Figure 2-6 shows the model responses for /æ/ (F0 = 100 Hz), /ɑ/ (F0 = 126 Hz) presented at 25 dB SPL. The peak in the PACF occurred at 10 ms, which corresponds to correct F0 estimation (= 100 Hz), as indicated by an arrow in PACF [Fig. 2-6(B)]. The percent ACF channels that showed a peak at 10 ms was 44%, which was marginally higher than the user-defined segregation parameter (43%). Thus, the model identification is based only using the *m1/m2* criterion. The model response was /æ/, /æ/, which was the first confusion (identified as

31%) in listeners' responses. Additionally, if *m1/m2* ratio was changed to 2.56, then the model was successful in predicting the correct answer.

Figure 2-7 shows the model responses for /æ/ (F0 = 100 Hz), /ɑ/ (F0 =126 Hz) presented at 65 dB SPL. There was a peak at 10 ms in the pitch region of the PACF [Fig. 2-7(D)] and thus the dominant F0 = 100 Hz was estimated correctly. The model does F0 segregation as 50% of the ACF channels showed a peak at 10 ms, which was less than the segregation parameter (52%). The ACF channels, that had a peak at 10 ms, were grouped together [Fig. 2-7(B)] whereas the remaining channels were grouped separately [Fig. 2-7(C)]. Two

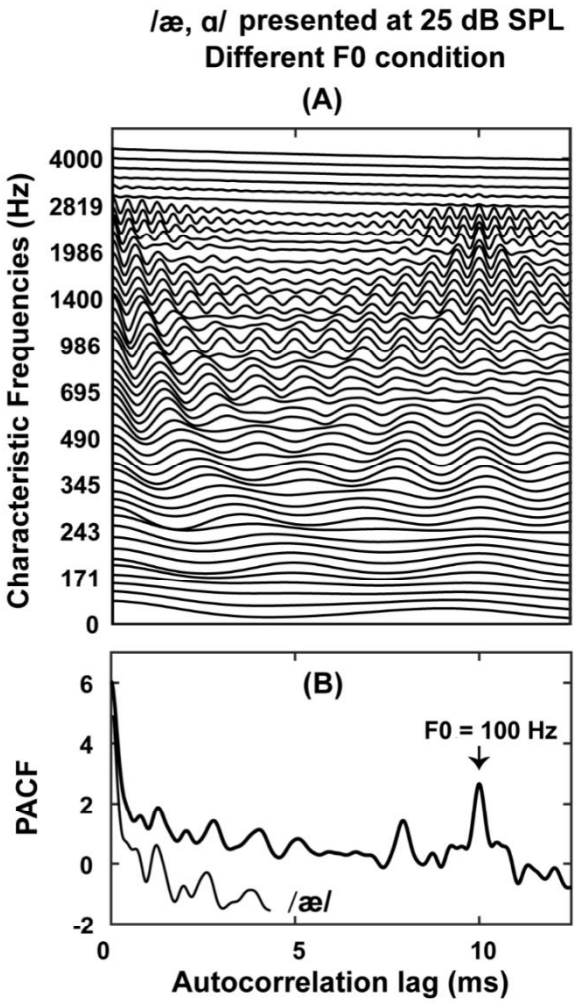

Figure 2-6 Model responses for /æ/ (F0 = 100 Hz), /ɑ/ (F0 = 126 Hz) presented at 25 dB SPL. The caption is similar to one of the columns in Figure 2-5.

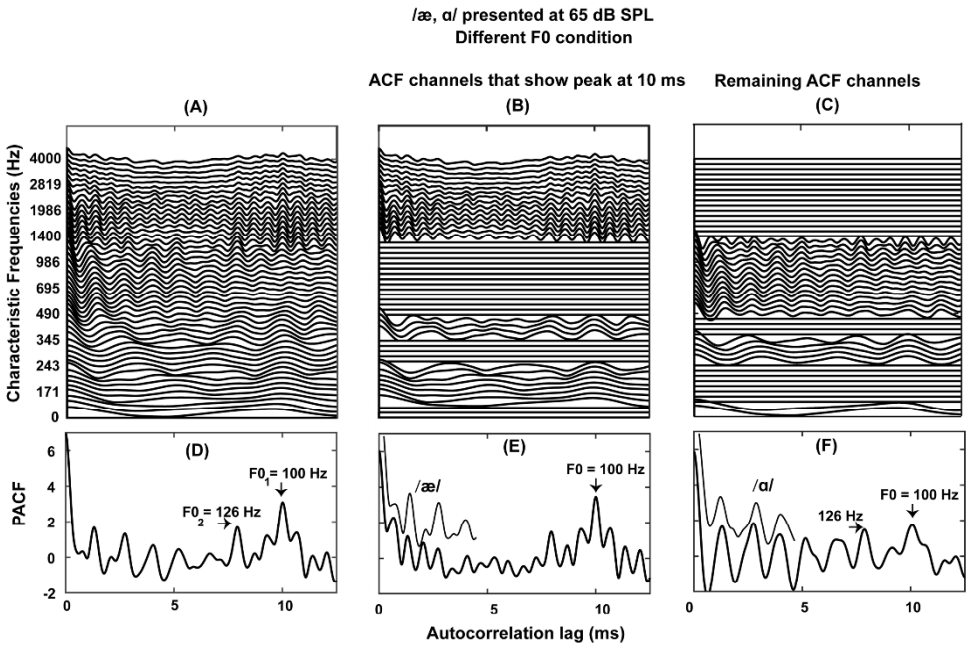

Figure 2-7 Model responses for /æ/ (F0 = 100 Hz), /ɑ/ (F0 = 126 Hz) presented at 65 dB SPL. The first column corresponds to the individual ACF channels from 100 different AN fibers. These channels are added together to obtain the pooled ACF (panel D). The two fundamental frequencies in the panel D, are denoted by $F0_1$ and $F0_2$. The estimated F0 is $F0_1$ (= 100 Hz). The second column shows only ACF channels that have peak at 10 ms (panel B) and the remaining channels are placed in the third column as shown in panel (C). The model responses are correct as shown in panels (E) and (F). Note that the timbre regions of the templates /æ/ and /ɑ/ (solid lines) are shown in panels (E) and (F) with an arbitrary vertical and horizontal offsets for clarity. For visualization purposes, only 50% channels are shown in the ACF plots.

segregated PACFs were computed from these groups. For vowel identification, the timbre region of the segregated PACF was compared with the previously stored templates of five different single vowels. The model identified /æ/ and /ɑ/ correctly [Figs. 2-7(E) and 2-7(F)]. The respective F0 of the vowel was correctly identified only in one case [i.e., F0 = 100 Hz, Fig. 2-7(E)] but in another case, the PACF indicated that there were two prominent peaks at 100 Hz, PACF amplitude = 1.74, and 126 Hz, with PACF amplitude = 1.49 [Fig. 2-7(F)]. Since only one of the F0s needs to be estimated correctly for vowel segregation using F0 difference (Meddis and Hewitt, 1992), the model was successful in identifying both the vowels.

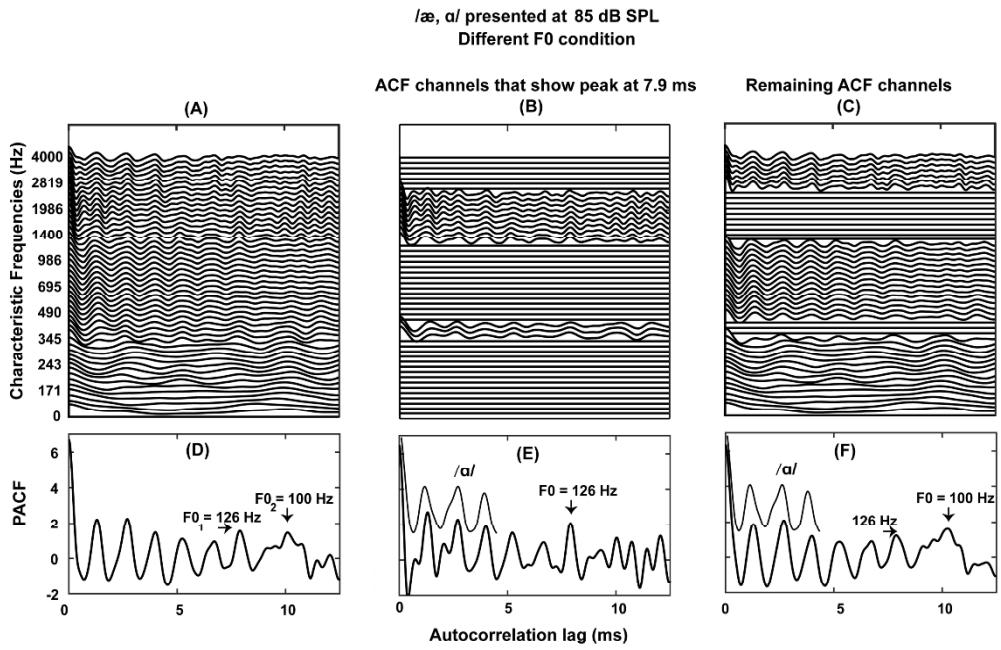

Figure 2-8 Model responses for /æ/ (F0 = 100 Hz), /ɑ/ (F0 = 126 Hz) presented at 85 dB SPL. The figure caption is similar to Figure 2-7. The estimated F01 = 126 Hz. The second column shows only ACF channels that have peak at 7.9 ms (panel B) and the remaining channels are placed in the third column as shown in panel (C). The model responses are shown in panels (E) and (F). The model response is correct for panel (E) but incorrect for panel (F). The timbre region of the template /ɑ/ (solid line) is shown in panels E and F.

Figure 2-8 shows the model responses for /æ/ (F0 = 100 Hz), /ɑ/ (F0 = 126 Hz) presented at 85 dB SPL. The dominant F0 was estimated correctly, as there was a peak at 7.9 ms (marginally higher than 10 ms) in the pitch region of the PACF [Fig. 2-8(D)]. Since only 24% of the ACF channels had a peak at 7.9 ms, F0 based vowel segregation was allowed prior to its identification. The model identified /ɑ/ with F0 = 126 Hz [Fig. 2-8(E)] correctly but incorrectly responded to /ɑ/ with F0 = 100 Hz [Fig. 8(F); note that the second peak in the PACF corresponds to 126 Hz]. The response of the model was /ɑ, ɑ/, which was the second confusion (identified as 11%) and the correct identification score was 55% observed in listeners' responses. At 85 dB SPL due to improper vowel segregation, the model failed to obtain the correct answer.

2.4 Discussion

2.4.1 Sensitivity of predicted concurrent-vowel identification scores to AN-model parameters

One recurrent question that arises in studies that use AN models (largely derived from animal studies) to predict human performance is whether the exact parameters used for cochlear frequency selectivity and spontaneous-rate distributions are appropriate for humans. Thus, this section investigates how changes in the AN-model parameters (e.g., tuning and SR) affect identification score for both vowels. There is a strong evidence that human tuning is sharper than animals (e.g., Joris et al., 2011; Shera et al., 2002), and thus we have directly altered the auditory-filter bandwidth at each CF to simulate an increase in tuning sharpness in the cat AN model (Zilany et al., 2014). To address the sensitivity issue associated with changes in SR, the default values for high, medium and low-SR fibers (i.e., 100, 4 and 0.1 spikes/sec) of the AN model (Zilany et al., 2014) were altered to 20, 2 and 0.01 spikes/sec, respectively. For both conditions (i.e., changes in tuning and SRs), it was found that the model scores can successfully capture the qualitative level-dependent effects for the same- and different-F0 conditions. The main requirement was that the pattern of level-dependent changes in vowel segregation for the different-F0 condition was preserved and also that the vowel segregation was zero at each level for the same-F0 condition [Fig. 2-3(B)]. More specifically, the pattern of vowel segregation with increasing level for the different-F0 condition has to be similar, although the absolute values could differ. Thus, the overall conclusions of this modeling study do not appear to be dependent on the specific parameter choices related frequency tuning and spontaneous-rate distributions. It is possible that a

better quantitative fit to the concurrent-vowel data could be obtained by altering the *m1/m2* value slightly.

2.4.2 Effect of *m1/m2* on identification of identical vs. different vowels

The selection of *m1/m2* will influence the model response (either identical or different vowels; Fig. 2-2) under the no-F0-segregation condition and will greatly affect the percent correct identification. Thus, in the current objective, *m1/m2* value was varied from 1.25 to 3, with increments of 0.25. It was found that *m1/m2* = 2.5 was successful in predicting the identification scores of both vowels [Fig. 2-3(A)] and F0 benefit [Fig. 2-3(D)] at least qualitatively, and one-vowel identification quantitatively [Fig. 2-3(C)]. Additionally, when averaged across vowel pairs and levels for the same-F0 condition, the percentage of model responses that were either correct or were a confusion observed in the data was 78%. For the different-F0 conditions, the model had only 45 vowel pairs (out of 150) across levels that went to a no-F0-segregation condition and the percent model response either for correct answer or observed confusion was 82%.

2.4.3 Selection of single-vowel templates for identification

The optimal single-vowel templates at each level (i.e., 5 vowel templates per level) were used for identifying the vowels. The model response could be influenced by the manner in which the single-vowel templates are generated. For example, when a single template for each vowel was generated regardless of level (i.e., only 5 vowel templates), it was found that model scores did not capture the patterns of level-dependent changes in identification scores for the same- and different-F0 conditions. Thus, this analysis further suggests that the

templates have to be different for each level due to the level-dependent changes in cochlear non-linearities (Ruggero et al., 1997).

2.4.4 Possible physiological mechanisms underlying the level-dependent changes in identification scores of concurrent vowels

The model scores of concurrent-vowel identification suggest that the level-dependent changes in phase-locking of AN fibers to vowel formants and F0s can account for trends in segregation and identification of concurrent vowels across levels and F0 conditions. Thus, the present predictions provide quantitative support for the qualitative suggestions from Chintanpalli et al. (2014). However, the model responses were lower across levels and F0 difference, as compared to identification data [compare Fig. 2-3(A) vs. Fig. 2-1], suggesting that other cues, apart from phase locking, could also be contributing. Rate-place cues could further improve the scores for low-to-mid levels, as it has been shown to encode vowel's formants (Sachs and Young, 1979) and F0s (Keilson et al., 1997). Other central auditory cues (e.g., cognition, attention) can also interact with these identification scores of both vowels. However, the modeling adopted in this study did not take into account the rate-place of the periphery and central auditory cues.

2.4.5 Conclusion

The present computational modeling study predicted the level-dependent changes in identification scores of both vowels for the cases with and without F0 difference between the vowels being presented. The model was developed by cascading the physiologically realistic AN model with a modified version of Meddis and Hewitt F0 based segregation algorithm (for complete block diagram,

Fig. 2-2). The model was tested against the identification score of both vowels and identification score of one vowel of the pair as a function of level for same- and different-F0 conditions. The segregation parameter of Meddis and Hewitt (1992) modeling was controlled by the actual F0 benefit across levels. The results from the modeling revealed that vowel segregation (either using the F0 difference or formant difference) based on the temporal coding cues of AN fibers can qualitatively explain the level-dependent changes in identification scores of both vowels with and without F0 difference [Fig. 2-3(A)]. Additionally, the model was successful in capturing the identification of one vowel quantitatively [Fig. 2-3(C)] and F0 benefit qualitatively [Fig. 2-3(D)].

.