3 Modeling concurrent vowel identification for shorter durations[3,4]

3.1 Introduction

In a cocktail party scenario, the YNH listeners have difficulty in understanding the target speech, when stimulus duration is reduced. For example, they have more difficulty in pitch-matching and F0-discrimination experiments (e.g., with complex tones) as stimulus duration is reduced (Gockel et al., 2007; Moore et al., 1986). Similarly, behavioral studies on concurrent vowels indicate that the effect of F0 difference was reduced at 50 ms compared to 200 ms duration. Two studies showed a reduced F0 benefit (Assmann and Summerfield, 1994; McKeown and Patterson, 1995), while others showed no F0 benefit (Assmann and Summerfield, 1990; Culling and Darwin, 1993).

Two possible neural mechanisms that may account for the effect of F0 difference on the percent identification of concurrent vowels: (1) a temporal envelope cue, due to waveform interactions of the two vowels, especially at lower F0 differences (e.g., 1.5 - 3 Hz), and (2) an F0-guided vowel segregation cue. It has been speculated that the temporal envelope cue may account for identification at lower F0 differences for concurrent vowels, regardless of its duration (i.e., 1000 ms as in Culling and Darwin, 1994; 50 ms as in Assmann and Summerfield, 1994). Additionally, Culling and Darwin (1994) had used a neural network model to successfully predict the identification scores based on temporal

[3] This work was published in the Speech Communication Journal, vol. 125, 1 – 6, 2020. doi:10.1016/j.specom.2020.09.007

[4] The preliminary portion of this work was published in the 42nd meeting of Association of Research in Otolaryngology in Baltimore, USA, 2019.

envelope cue at lower F0 differences. However, Meddis and Hewitt (1992) model was successful in capturing the effect of F0 difference on concurrent vowels, with 200 ms duration, solely based on an F0-guided segregation cue. They utilized an AN model with an F0-guided segregation algorithm to predict the percent correct of both vowels as a function of F0 difference. Later on, the similar F0-guided segregation algorithm but with a more recent and physiologically realistic AN model was used to capture the effect of F0 difference on concurrent-vowel scores observed in normal-hearing listeners, either at a particular level (Chintanpalli and Heinz, 2013) or across levels (Settibhaktini and Chintanpalli, 2018). The studies mentioned above suggest that both waveform interactions and F0-guided segregation cues are available for normal-hearing listeners for identifying the concurrent vowels.

Based on an F0-guided segregation cue, the second modeling study attempts to test the hypothesis that a limited ability to utilize this cue could contribute to reduced concurrent-vowel scores for shorter durations. Here, we predicted concurrent-vowel identification scores across F0 differences for longer and shorter durations. We used population responses from an AN model (Zilany et al., 2014) and a modified version of the Meddis and Hewitt (1992) F0-guided segregation algorithm. Additionally, differential reductions in the ability to avail the F0-guided segregation cue might explain individual differences in F0 benefit observed in concurrent-vowel data.

3.2 Methods

3.2.1 Stimuli

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). Each vowel duration was either 200 or 50 ms, including 15 ms raised-cosine rise and fall ramps. The center frequencies and bandwidths of the five formant (F1 – F5) filters used to generate these vowels are shown in Table 2-1 and were similar to the previous concurrent vowel studies (e.g., Assmann and Summerfield, 1994; Chintanpalli et al., 2014; Summers and Leek, 1998).

A concurrent vowel stimulus was generated by summing a pair of single vowels. In each concurrent-vowel pair, one vowel's F0 was always fixed at 100 Hz while the other's F0 was either 100, 101.5, 103, 106, 112 or 126 Hz. For each condition, there were 25 concurrent vowel pairs. To maintain an equal number of vowel pairs, the 0-Hz F0 difference condition had five identical vowel pairs and ten different vowel pairs, where the latter was repeated twice. There were 150 concurrent vowels (25 pairs x 6 F0 differences) at each duration. Overall, a total of 300 concurrent vowels (150 pairs x 2 durations) were used. Each concurrent vowel was presented to the AN model as a sound waveform (in units of Pascals) presented at ~68 dB SPL, as this level is a representative of the average of the range of the levels (63 – 74 dB SPL) used in Assmann and Summerfield (1994). The individual vowels were presented at 65 dB SPL. Compared to the stimuli section for first objective (section 2.3.1), the vowel duration was varied at six F0 difference conditions for a fixed level (65 dB SPL).

3.2.2 Computational auditory-nerve model

The second objective in this dissertation utilized the same AN model (section 2.3.2) the was used for the first objective. The output of the model was a time-varying DR obtained for a single AN fiber tuned to a particular CF. These DRs were predicted for 100 different logarithmically spaced CFs, ranging from 250 to 4000 Hz. The population DR for a specific CF was the weighted sum of DRs as per the typical SR distribution of AN fibers (HSR = 0.61, MSR = 0.23 and LSR = 0.16; Liberman, 1978).

3.2.3 F0 guided segregation algorithm

The same F0-guided segregation algorithm, that was utilized in the first objective (section 2.3.2.2) was used in the current study. Only difference was that the CF-dependent time constant values were altered in order to account for the duration effects. In this study, Here, the time constant was varied with CF ($\Delta\tau$ = 80 ms for 250 ≤ CF < 440 Hz; $\Delta\tau$ = 66 ms for 440 ≤ CF < 880 Hz; $\Delta\tau$ = 60 ms for 880 ≤ CF < 1320 Hz; $\Delta\tau$ = 59 ms for CF ≥ 1320 Hz).

3.3 Results

Figures 3-1(A) and 3-1(B) show mean values of the normalized pooled-ACF for the two (actual) F0s of the 25 vowel pairs for various F0 difference conditions as a function of vowel duration. The actual F0s are the true (or known) F0s of the vowel pair. Compared with other F0 difference conditions, the salience (or strength) of the normalized pooled-ACF was higher for the 0-Hz F0 difference due to large F0 energy concentrated near 10 ms. As F0 difference was increased, the F0 energy was shared between two F0s of the vowel pair and thus resulted in a reduction of the salience at each F0 [Figs. 3-1(A) and 3-1(B)]. These

findings suggest that the salience of the two F0s decrease with an increasing F0 difference, regardless of the duration. For F0 differences (≥ 6 Hz), the two distinct F0 peaks were observed in the individual PACFs across durations. Under these conditions, the salience of the normalized pooled-ACF was larger for varying F0 than the fixed F0 of 100 Hz [compare Figs. 3-1(A) with 3-1(B)]. The varying (or higher) F0 will have a larger harmonic energy near the vowel's formant frequency. However, for F0 = 100 Hz, the energy will be shared between the two harmonics around the formant frequency. Thus, a large harmonic energy will elicit a strong response to a higher F0 after auditory filtering, as reflected in the pooled-ACF. This observation is specific to the stimulus set used in this study and other previous studies (Assmann and Summerfield, 1994; Chintanpalli et al., 2014; Summers and Leek, 1998). Consistent with the acoustic representations, the salience of each of the F0s was reduced with decreasing vowel duration for all F0 differences [Figs. 3-1(A) and 3-1(B)]. This indicates that shorter durations reduce the representations of both F0s.

To account for the F0 difference effect on concurrent-vowel identification, the Meddis and Hewitt (1992) F0-guided segregation algorithm requires at least one F0 to be estimated correctly for vowel segregation and correct identification of both vowels. Figure. 3-1(C) shows mean values of the normalized pooled-ACF for the correctly estimated dominant F0 of the 25 vowel pairs across F0 differences as a function of vowel duration. As expected, the salience of the dominant F0 was predicted to be reduced with decreasing vowel duration.

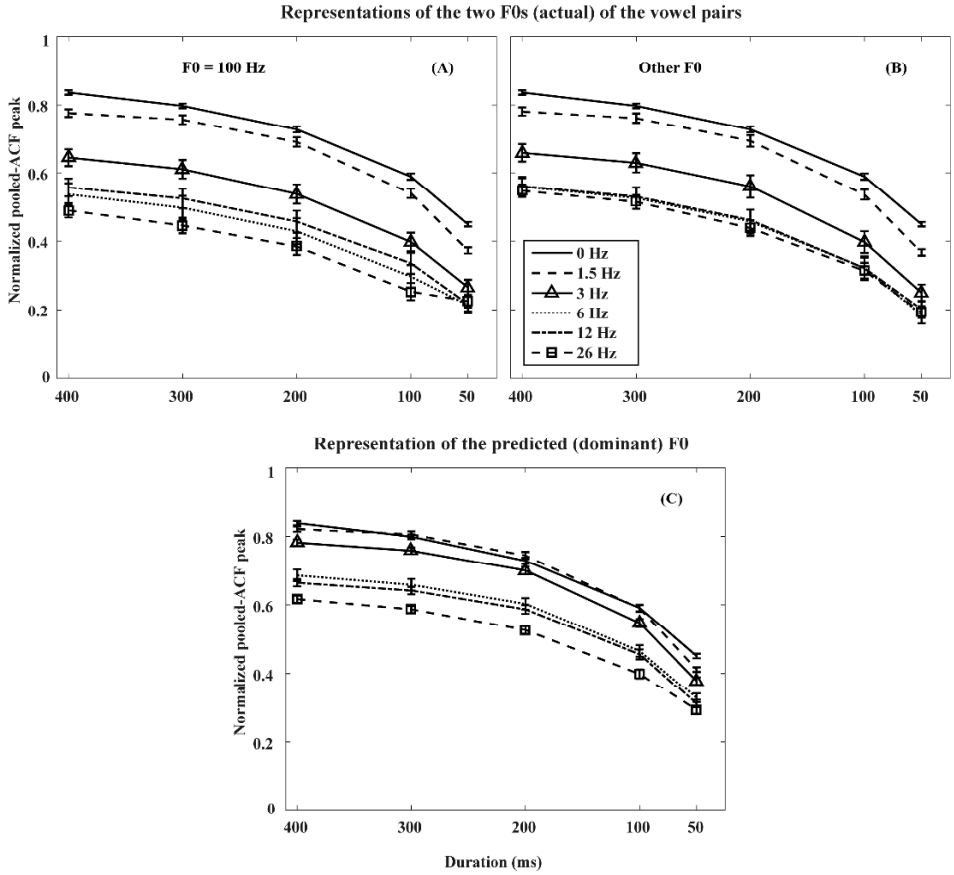

Figure 3-1 Effect of vowel-pair duration on mean normalized pooled-ACF values evaluated at the periods associated with (A) F0 = 100 Hz, (B) other F0 of the pair, and (C) the correctly predicted dominant F0 in the Meddis and Hewitt (1992) segregation algorithm. Note that dominant F0 was considered correct if it fell within ± 1 Hz of the true F0, with the 10% wrongly identified conditions (out of 750 vowel pairs across durations) not included in this analysis. In some cases, the normalized pooled-ACFs were higher in Figure. 1(C) than Figures. 1(A) and 1(B), as the pooled-ACF for the dominant F0 can occur anywhere within ±1 Hz window of the true F0 (either 100 or varying F0). For each F0 difference, the pooled-ACF was normalized by the maximum across durations. Legend indicates F0 difference conditions. Means: across all 25 vowel pairs. Error bars: ± 1 SEM.

Figures 3-2(A) and 3-2(B) show the model percent correct identification scores of both vowels and its corresponding percent segregation across F0 differences for 200 ms duration. Additionally, the F0-segregation parameter and identification parameter (*m1/m2*) criterion used were 85% and 1.75, respectively. Consistent with other modeling studies (Chintanpalli and Heinz, 2013; Meddis and Hewitt, 1992), the predicted scores (solid line) increased with F0 difference and then asymptoted at 6-Hz F0 difference [Fig. 3-2(A)]. The pattern of identification score as a function of F0 difference is qualitatively similar to

Assmann and Summerfield (1994) concurrent vowel data for 200 ms (dashed line). Consistent with Chintanpalli and Heinz (2013), the ability to segregate two vowels improved with an increasing F0 difference and reached the maximum at higher-F0 difference [Fig. 3-2(B)].

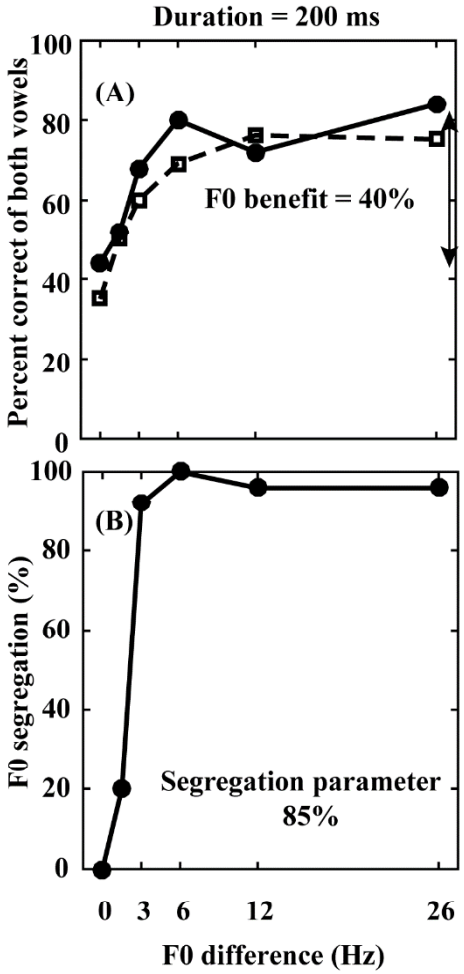

Figure 3-2 Predicted effects of F0 difference on concurrent vowel identification and segregation for 200 ms (solid lines) duration. (A) Percent identification scores of both vowels. (B) Percent F0 segregation of both vowels, computed as the proportion of vowel pairs (out of 25) in which the ACFs were segregated into two different sets. For comparison, Assmann and Summerfield's (1994) concurrent vowel data for 200 ms are shown in panel (A) as a dashed line. The predicted F0 benefit (shown by arrow) is shown in the top panel.

To assess whether the poor F0 representation for 50 ms [Fig. 3-1(C)] can solely account for the reduced concurrent-vowel identification scores, Figures. 3-3(A) and 3-3(B) show the model scores and its corresponding percent segregation across F0 differences using the same parameters used for 200 ms. The model scores for 50 ms [Fig. 3-3(A)] were reduced only at 3-Hz (from 68%

to 64%), 6-Hz (from 80% to 60%) and 12-Hz (from 72% to 64%) F0 difference conditions, relative to 200 ms duration. As the effect of poor F0 representation did not fully account for reduced concurrent-vowel scores across F0 differences, this suggest that the reduced segregation ability can be a contributing factor for 50 ms duration.

As the salience of the dominant F0 is not completely zero for 50 ms [Fig. 3-1(C)], there is a possibility that differences in the abilities to avail F0-guided vowel segregation cue might vary across subjects, resulting in individual differences in F0 benefit. This hypothesis was tested by altering only the F0-segregation parameter and keeping the remaining parameters the same (i.e., used for 200 ms). Figures 3-3(C) - 3-3(H) show the model identification of both vowels (solid lines in panels C, E, G) and its percent F0 segregation (panels D, F, H) for 50 ms duration, based on three different F0-segregation abilities (reduced from left to right). When the F0-segregation parameter was changed to 75%, the overall percent scores were reduced as a function of F0 difference [Figs. 3-3(C) and 3-3(D)]. Additionally, the overall scores were further reduced with a decrease in the segregation parameter to 55% [Figs. 3-3(E) and 3-3(F)]. In both the cases, the predicted F0 benefit was reduced (to 24% and 20%, respectively), in contrast to the 200 ms condition where the F0 benefit was 40%. When the model's percent F0-segregation was zero [i.e., for 15% or below, Figs. 3-3(G) and 3-3(H)], the scores were relatively constant and the F0 benefit was zero, indicating that the availability of the F0-guided segregation cue is completely lost. Among these, 75% segregation parameter model scores matched qualitatively with the concurrent vowel data [Fig. 3-3(C)]. The model

fitted the lower scores and variable F0 benefit across subjects, by varying the ability to utilize an F0-guided segregation cue.

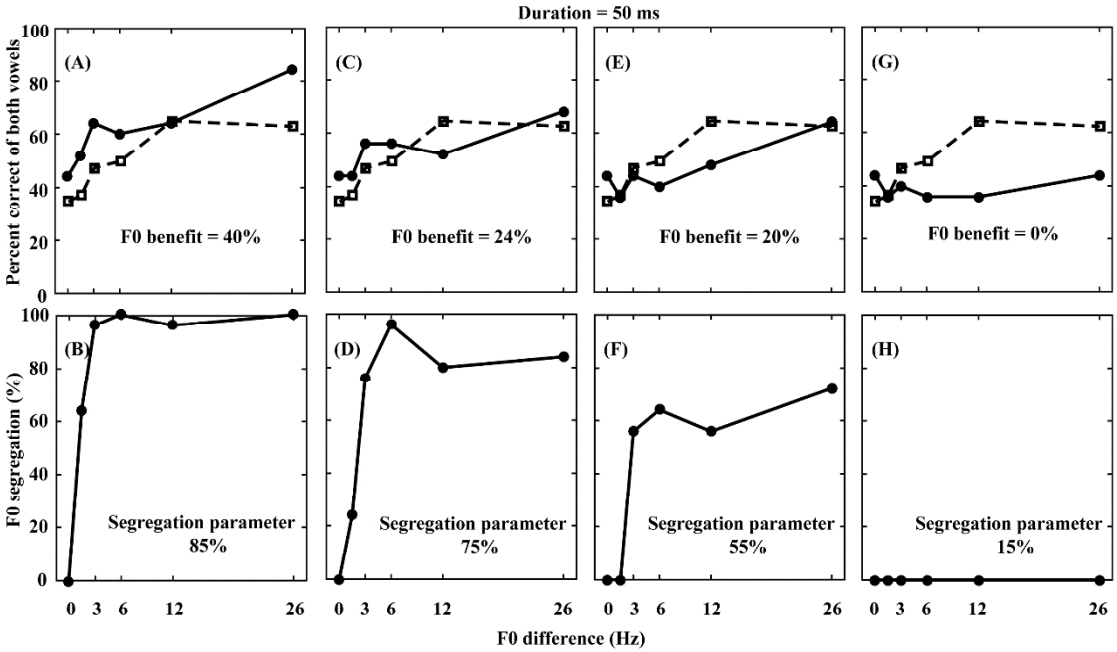

Figure 3-3 Predicted effects of F0 difference on concurrent vowel identification and segregation for 50 ms durations (solid lines). Top row: Percent identification scores of both vowels. Bottom row: Percent F0 segregation. Predictions for the 50 ms duration are shown for four different model segregation abilities, with the segregation parameter equal to 85%, 75%, 55% and 15%, respectively. For comparison, Assmann and Summerfield (1994) concurrent vowel data for 50 ms are shown in panels (A, C, E, G) as dashed lines. The predicted F0 benefit is shown in the top panels.

Figure 3-4 shows the model percent correct identification scores of one vowel of the pair as a function of F0 difference for 200 ms and 50 ms durations. For 200 ms [Fig. 3-4(A)], the score of one vowel was 100% and was independent of F0 difference. Comparison of Figures 3-2(A) (solid line) and 3-4(A) suggest that F0 difference is a primary important factor for identifying the second vowel of the pair. This finding is consistent with the concurrent vowel data (Chintanpalli et al., 2016; Chintanpalli and Heinz, 2013) and modeling predictions (Chintanpalli and Heinz, 2013; Settibhaktini and Chintanpalli, 2018). Similarly, the same observation was noticed for 50 ms, as the model scores of one vowel were 100% across F0 differences, regardless of the segregation parameter used in this study

[compare solid line of Fig. 3-3(C) with Fig. 3-4(B)]. When compared across durations (50 ms vs. 200 ms), the identification scores of the second vowel were improved with an increasing duration [compare solid line of Fig. 3-2(A) with Fig. 3-4(A) for 200 ms and solid line of Fig. 3-3(C) with Fig. 3-4(B) for 50 ms]. This modeling observation is consistent with McKeown and Patterson (1995) concurrent vowel data, which showed that improvement in identification scores of both vowels (with an increasing duration) is primarily due to the identification of the second vowel.

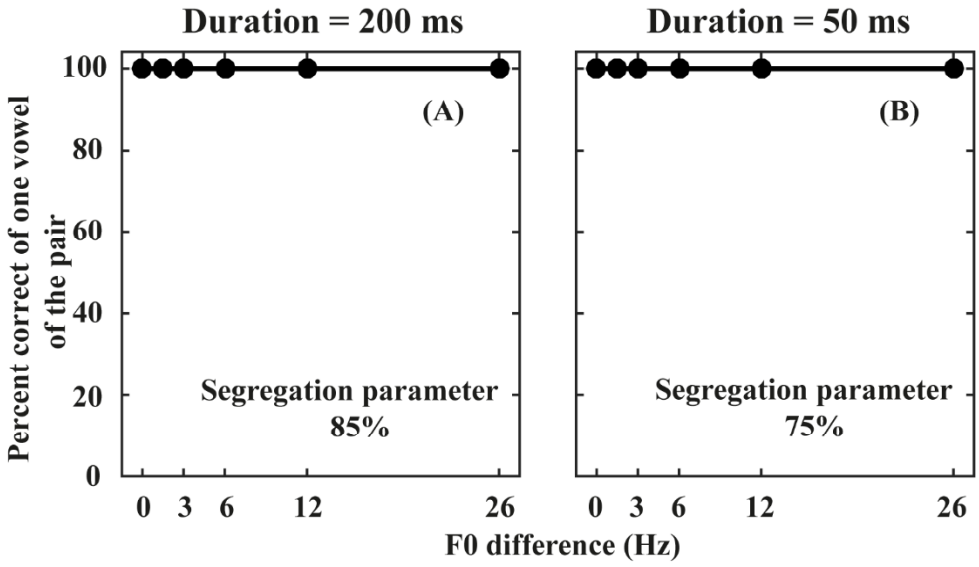

Figure 3-4 Predicted effects of F0 difference on identification scores of one vowel of the pair for two different vowel durations. (A) 200 ms, (B) 50 ms. Note that for 50 ms, the one vowel correct is always 100% regardless of the segregation parameter. For illustration purposes, only the segregation parameter equal to 75% is shown in panel B.

3.4 Discussion

The modeling predictions shown here are based on temporal responses of the AN fibers (Zilany et al., 2014) and a modified version of the Meddis and Hewitt (1992) F0-guided segregation algorithm. The F0 representations of the vowel pairs were reduced with decreasing duration (Fig. 3-1). For 200 ms duration [Figs. 3-2(A) and 3-2(B)], the model scores of both vowels across F0

differences were qualitatively similar to Assmann and Summerfield (1994) concurrent vowel data. By limiting the ability to avail the F0-guided vowel segregation cue for 50 ms, the model was successful in capturing the lower F0 difference effect. Additionally, the variable F0 benefit across normal-hearing listeners (Assmann and Summerfield, 1990, 1994; Culling and Darwin, 1993; McKeown and Patterson, 1995) can be predicted by varying the differential ability to avail the F0-guided segregation cue [Figs. 3-3(C) – 3-3(H)]. The model was also successful in predicting the improvement in identification of the second vowel, when vowel duration was increased from 50 ms to 200 ms (consistent with McKeown and Patterson (1995) concurrent vowel data).

Different parameter sets of the F0-guided segregation algorithm (i.e., ACF time constant, F0 segregation parameter, *m1/m2* criterion) can also be used, other than those in the current study. These parameters could also vary across sound levels. However, as long as, these segregation parameters can capture the effect of F0 difference on concurrent-vowel scores for 200 ms duration, then by varying only the F0-segregation parameter for 50 ms, it is expected that the main conclusion of the study will not be altered.

The two distinct F0s of the concurrent vowel pair were predicted correctly in the current study, when the F0 difference was 6 Hz (i.e., 1 semitone) or higher. This finding is consistent with the concurrent harmonic tones, separated either by one or four semitone (Larsen et al., 2008). However, these two studies are in contrast with Assmann and Paschall (1998) concurrent vowel data, where their listeners heard only single F0 below four semitones. Larsen et al. (2008) speculated that an additional behavioral data on F0 identification with the

concurrent complex tones are required to understand the possible reasons underlying this discrepancy. With respect to the current study, a more direct comparison would be to collect an additional behavioral data by asking the listeners to identify directly F0s of the concurrent vowels, rather than matching individually with the single harmonic tone complex as done in Assmann and Paschall (1998). This discrepancy can be addressed once the relevant behavioral data is collected.

The current modeling can be extended to predict the effects of SNHL on identifying the concurrent vowels with shorter durations. Zilany et al. (2014) model can also obtain the AN responses for hearing-impaired fibers to concurrent vowels. Different degrees of SNHL on concurrent vowel scores across F0 differences for 50 ms can be obtained, using the same parameters of the F0-guided segregation algorithm used in this study. These predictions may help us design an experimental paradigm (e.g., a suitable sound level) for hearing-impaired listeners, for which behavioral data is currently unavailable in the literature.

The current modeling investigated the effect of shorter durations on the role of an F0-guided segregation cue on reduced identification scores across F0 differences. The role of waveform interactions on reduced identification scores is yet to be addressed, as mentioned by Assmann and Summerfield (1994). A possible future work could be to collect additional behavioral data on concurrent vowels with different segments for 200 ms and 50 ms, using the same set of normal-hearing listeners. This experimental paradigm was used by Assmann and Summerfield (1994) but only for 50 ms. At a lower F0 difference, if the

difference in identification scores between two segments of 200 ms is greater than 50 ms segments, then there may be a decline in waveform interactions with decreasing duration. Alternatively, the same experimental paradigm or neural network model of Culling and Darwin (1994) can be extended for 200 ms and 50 ms durations, to evaluate the role of waveform interactions.

3.5 Conclusion

This current modeling study aimed to understand a possible underlying cause for reduced concurrent vowel scores for 50 ms in normal-hearing listeners, compared to scores for 200 ms, across F0 differences. We hypothesized that a limited ability to avail an F0-guided segregation cue could contribute to the reduced concurrent vowel scores. Furthermore, a differential reduction in the ability to utilize this cue might explain F0-benefit variability across subjects. To test this hypothesis, a physiologically realistic AN model (Zilany et al., 2014) with a modified version of Meddis and Hewitt (1992) F0-guided segregation algorithm was employed to predict the concurrent vowel scores. When the ability to use an F0-guided segregation cue of the Meddis and Hewitt (1992) algorithm was limited, the concurrent vowel scores across F0 differences were reduced qualitatively for 50 ms [Fig. 3-3(C)], when compared to 200 ms [Fig. 3-2(A)]. The differential reduction in this cue resulted in a variable F0 benefit [Figs. 3-3(E) and 3-3(G)]. Thus, these modeling predictions confirm our hypothesis that the inability to use an F0-guided segregation cue contributes to the reduced concurrent vowel scores for 50 ms (shorter) duration.