Experience & The Perception of Coarticulated Speech

Andrea Pearman & Susagna Tubau
Universitat Autònoma de Barcelona, Spain
E-mail: apearman225@yahoo.es, stubau@hotmail.com

ABSTRACT

This paper investigates whether perception of coarticulatory effects, specifically the spectral effects exerted on and by adjacent segments, is influenced by one's L1. Four continua, [ʃ̫u:ʃ̫] – [ʃ̫i:ʃ̫],[ʃ̫ɑʃ̫] – [ʃ̫ɛʃ̫] and their isolated vowel counterparts, were presented as native language sounds to North American English and Catalan speakers. Lip rounding, associated with a lowering of the second and third formants, is a typical feature of alveolo-palatal fricatives in English, but not in Catalan. Thus, if listeners' identification of the speech signal is in fact influenced by their L1, we expect that Catalans will perceive more back vowels in a lip rounded [ʃVʃ] context than Americans, who should perceive more front vowels. Additionally, Americans are expected to show similar identification curves for the two conditions, while Catalans should show different curves: reporting more back vowels in context than in isolation. Resulting trends support these predictions.

1. INTRODUCTION

Compensation for coarticulatory variation has been evidenced in a number of studies, e.g. [1, 2, 3, 4, 5, 8, 10]. Listeners are known to normalize in cases where variation can be related to the conditioning context; however, they are known to fail to normalize in cases where the variation cannot be related to the context [1, 5, 10]. Unable to factor out the effects of coarticulation, they interpret the signal at face value. Data concerning the effects of nasalization on vowels show that either phonetically or phonologically inappropriate nasalization may influence the perception of vowel height [1]. Subjects identify nasalized high vowels as lower vowels when the nasal effect is insufficient or excessive or when a nasal consonant is not immediately adjacent to the vowel, in the case of languages lacking distinctive nasal vowels. Therefore, speakers interpret the various spectro-temporal changes in the acoustic signal in terms of the criterial cues used in the L1 to indicate particular sounds [6].

Cross-linguistic differences in secondary articulations and in the extent and magnitude of coarticulation have been reported previously, e.g. [9]. In the present study, we aim to demonstrate that listeners' ability to compensate for coarticulatory effects depends on their exposure to or familiarity with such effects. If the listener is unfamiliar with other language or dialect-specific effects, these will not be compensated for and will be processed at face value, *i.e.* as intended variation. We specifically focus on differences of alveolo-palatal /ʃ/ in American English and Catalan. In American English, /ʃ/ typically involves a measure of lip rounding, as do /ʒ/, /tʃ/, and /dʒ/, which is associated with the lowering of F2 and F3 in coarticulated vowels. In Catalan, /ʃ/ does not involve lip rounding [7]. Therefore, if listeners' perception of the speech signal is influenced by their L1, when presented with back to front vowels in isolation as well as in a rounded alveolo-palatal context, [ʃ̫], we predict that: 1) Catalans will perceive fewer front, *i.e.* more back, vowels in context vis-à-vis American speakers, because they will not relate the lowering of F2 and F3 induced by labialization to its physical source, [ʃ̫]. Rather, Catalans will perceive the lower formants as a property of the vowel, producing a face value interpretation: a more 'back' vowel. Americans, on the other hand, will perceive more front vowels, in comparison to Catalans, since they are expected to relate the effects of labialization to the source and score them out, leaving the interpretation of the vowel unaffected. 2) Identification functions of vowels in context and in isolation for Catalan subjects should be different. Vowels in context are expected to be perceived as more 'back,' due to the lack of correction for unfamiliar coarticulatory effects. For Americans, however, the two functions are expected to be roughly the same. Due to the Americans' experience, vowels in context should be perceived similarly to vowels in isolation.

2. METHOD

Two back to front vowel continua in a lip rounded alveolo-palatal fricative context, [ʃ̫u:ʃ̫] – [ʃ̫i:ʃ̫] and [ʃ̫ɑʃ̫] – [ʃ̫ɛʃ̫], were synthesized using the ASL software, Model 4300. A native speaker of General American English recorded the ends of the continua. Each vowel's first three formants were measured and 8 separate stimuli were calculated in step changes, from token 1 [ʃ̫u:ʃ̫] to token 8 [ʃ̫i:ʃ̫] and similarly for 1 [ʃ̫ɑʃ̫] to 8 [ʃ̫ɛʃ̫]. For each, the target formant frequencies were kept constant for 50 ms. at the center of the vowel. Formant frequency values were raised by 50 Hz at the onset and the offset of the vowel and interpolated from the steady-state center portion to approximate naturally-occurring transitions. For each stimulus the length of the vowel was 200 ms. Two continua

ISBN 1-876346-48-5 © 2003 UAB

of vowels in isolation, [u:] – [i:] and [ɑ] – [ɛ], were also created by eliminating the fricative segments and the corresponding formant transitions for each synthesized token.

The experiment had 2 conditions (1) vowels in isolation, the control tokens and (2) vowels in context, the test tokens. The control tokens were used to obtain the reference identification functions of the vowel continua for each group. For [ʃu:ʃ] – [ʃi:ʃ] and [u:] – [i:], two blocks of 24 stimuli for each continuum were created with tokens occurring three times per block. Since this listening test proved to be too long, for [ʃɑʃ] – [ʃɛʃ] and [ɑ] – [ɛ] two blocks of 18 stimuli with tokens occurring at a 1-2-3-3-3-3-2-1 frequency were made, focusing more on those stimuli falling near the phoneme boundary.

Stimuli were presented in Power Point on a Sony portable computer, model PCG-GRZ530 Pentium IV. The subjects were tested in a quiet room and listened to the stimuli over headphones. Tokens were presented as native language sounds. Subjects were asked to label each as either shoosh/sheesh, oo/ee, or shosh/shesh, o/e (xuix/xix, u/i or xoix/xeix; o/e, adapted for Catalan speakers). Clear examples of the ends of the continua were provided at the beginning of each block. Subjects clicked on an icon prompting the sound when ready and recorded responses by hand, advancing through the presentation at their own pace. The duration of the test was approximately 30 minutes for [ʃu:ʃ] – [ʃi:ʃ] and [u:] – [i:] and 15 minutes for [ʃɑʃ] – [ʃɛʃ] and [ɑ] – [ɛ]. Subjects took a brief pause after each block.

In total, each listener gave 96 forced-choice responses (2 blocks of 24 stimuli for each of the 2 continua) for the [u:] – [i:] vowel set and 36 responses (1 block of 18 stimuli for each of the 2 continua) for the [ɑ] – [ɛ] set. Chi-square tests with a correction factor were performed on the data to see whether experience with the resulting coarticulatory effects of [ʃ] influences the identification of back *vs* front vowels.

For the [u:] – [i:] continua, 20 subjects participated in the experiment: 10 North American English speakers and 10 central-dialect Catalan speakers, providing a total of 120 observations per stimulus. For [ɑ] – [ɛ], 16 Americans and 16 Catalans were recruited, yielding 32, 64, and 96 observations per stimulus depending on the frequency.

3. RESULTS

The percentage of front vowel, [i:] and [ɛ], identification, both in isolation and in context, was calculated for each stimulus, shown in Figures 1 to 4. Chi-square tests on intermediate stimuli, 3 to 6, were performed with experience with the coarticulatory effects of [ʃ] (American *vs* Catalan) as the independent variable and perception of back *vs* front vowels (/u:/ *vs* /i:/, /ɑ/ *vs* /ɛ/) as the dependent variable.

For each figure, the ordinate represents the percentage of identification of the front vowel in each vowel set: [i:], for [ʃu:ʃ] – [ʃi:ʃ]/[u:] – [i:],and [ɛ], for [ʃɑʃ] – [ʃɛʃ]/[ɑ] – [ɛ]. Stimuli numbers are plotted along the abscissa from 1, [u:]/[ɑ], and 8, [i:]/[ɛ].

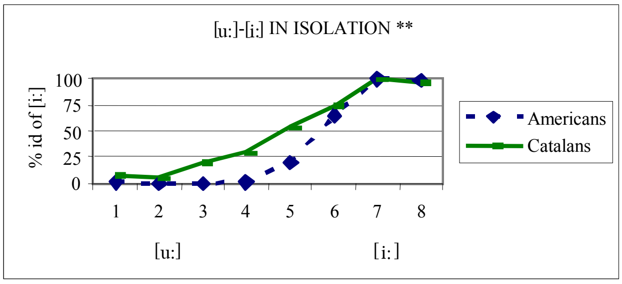

Figure 1: Identification functions of [i:] in isolation, $p<0.01$.

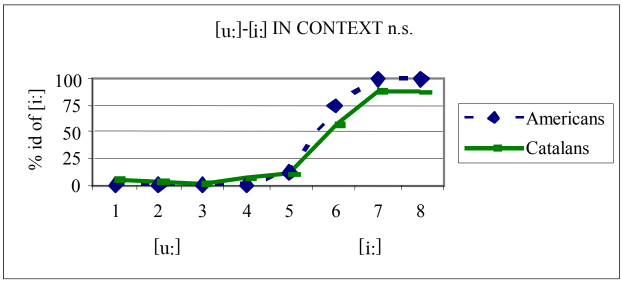

Figure 2: Identification functions of [i:] in context, $p>0.05$.

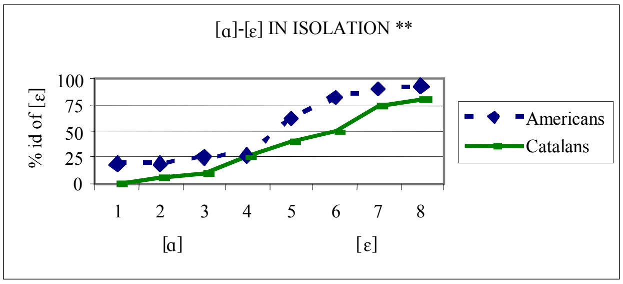

Figure 3: Identification functions of [ɛ] in isolation, p<0.01.

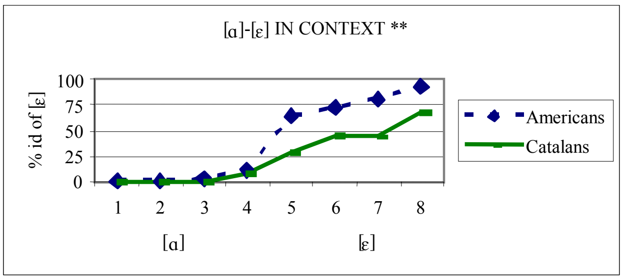

Figure 4: Identification functions of [ɛ] in context, $p<0.01$.

Figure 1 illustrates the percentage of Americans' and Catalans' front vowel identification in isolation, [i:], for the

ISBN 1-876346-48-5

[u:] – [i:] continuum. The major points of difference are found in the intermediate stimuli. For Americans, the phoneme boundary, or 50% cross-over point, falls between stimuli 5 and 6; while, for Catalans, it falls between stimuli 4 and 5. Thus, Catalans began to hear front vowels earlier than Americans in this continuum. Differences between the two functions are significant, $\chi^2_{(1)} = 27.424$, $p < 0.01$.

The identification curves for [u:] – [i:] in context are illustrated in Figure 2. The two functions are extremely similar. Both Americans and Catalans begin to consistently hear a greater percentage of [i:] after stimulus 5. For Americans, the phoneme boundary in context and in isolation is crossed between the same stimuli; for Catalans, the boundary is crossed one step later in context than in isolation. The results for the two curves in context are not significant, $\chi^2_{(1)} = 0.465$, $p > 0.05$.

The percentage of identification of front vowels in isolation, [ɛ], for Americans and Catalans for the [ɑ] – [ɛ] continuum is seen in Figure 3. Whereas for the isolated [u:] – [i:] continuum, Catalans begin to hear front vowels earlier than Americans; in the [ɑ] – [ɛ] in isolation continuum, Americans begin to hear front vowels significantly earlier than Catalans. The phoneme boundary for this continuum is crossed between stimuli 4 and 5 for Americans and not until between stimuli 6 and 7 for Catalans. The results of the chi-square test, $\chi^2_{(1)} = 12.828$, $p < 0.01$, are significant.

The percent identification for the [ɑ] – [ɛ] continuum in context for Americans and Catalans is shown in Figure 4. Much like the curves in context for [u:] – [i:], Americans and Catalans seem to perceive the early steps of the continuum similarly. At stimulus 4 however, the curves begin to diverge radically. Americans begin hearing front vowels between stimuli 4 and 5, the same point at which they began hearing front vowels for this vowel set in isolation; Catalans, however, do not begin hearing front vowels until much later, between stimuli 7 and 8. Like the Catalans' performance for [u:] – [i:] identification, Catalans begin hearing front vowels for [ɑ] – [ɛ] in context, a full step later than they do in isolation. The differences between the curves are significant, $\chi^2_{(1)} = 12.113$, $p < 0.01$.

4. DISCUSSION

Based on cross-linguistic differences in the realization of /ʃ/ in English and Catalan and the level of familiarity with the acoustic consequences of labialized alveolo-palatal fricatives, Catalans were expected to identify fewer front vowels in an alveolo-palatal context than Americans, who were predicted to identify more front vowels. Catalans were expected to perceive the physical consequences of lip rounding, F2 and F3 lowering, as a property associated with the vowel and not the consonant and thus perceive fewer front vowels in context, vis-à-vis Americans. Americans, in turn, were expected to identify the acoustic consequences of a lower F2 and F3 as a feature of the labialized fricative, score out the effects, and perceive more front vowels in context in comparison to Catalans. Results are shown in Figures 2 and 4. The percentage of front vowels identified for the [u:] – [i:] in context continuum, is represented in Figure 2. Though both Catalans and Americans begin to hear front vowels at the same point, between stimuli 5 and 6, Americans actually identify a slightly higher percentage of front vowels: 11.67% in stimulus 5 and 75% in stimulus 6 compared to 10% and 56.67% respectively for the Catalans. Chi-square results on stimuli 3 to 6 however are not significant, $\chi^2_{(1)} = 0.465$, $p > 0.05$. Regarding the [ɑ] – [ɛ] in context continuum, Figure 4, Catalan speakers clearly heard fewer front vowels than Americans, as predicted, since the back to front cross-over does not happen until quite late in the continuum, after stimulus 7. Americans began hearing front vowels a full three steps earlier, between stimuli 4 and 5. The differences for the two curves are significant, $\chi^2_{(1)} = 12.113$, $p<0.01$.

According to our second prediction, the identification functions of vowels in isolation *vs* vowels in context for Catalans should be significantly different. As Catalans do not have experience with F2 and F3 lowering as a result of lip rounded alveolo-palatal fricatives in their native language, we predicted that they would perceive the lower formant frequencies as a property of the vowel, and provide a different interpretation of vowels in context *vs* vowels in isolation. Conversely, we expected that the two functions, in isolation and in context, for Americans would be roughly the same. Vowels in context were expected to be perceived similar to vowels in isolation, since the F2 and F3 lowering induced by labialized /ʃ/ was expected to be recognized immediately due to experience, its effects being perceptually factored out. Comparing Figures 1 and 2, [u:] – [i:] in isolation and in context, for Catalans (the solid lines), the two functions are quite different; for the intermediate stimuli, great disparity is seen. Catalans begin hearing front vowels later in context, after stimulus 5, than they do in isolation, after stimulus 4. We associate the significant differences between the vowel in context and the vowel in isolation continua, $\chi^2_{(1)} = 35.825$, $p < 0.01$, as being attributable to the physical and perceptual effects induced by lip rounding. The physical effects, F2 and F3 lowering, actually a consequence of a feature of the fricative, are perceptually integrated as a property of the vowel, which is thus perceived as lower. For Americans (shown as the dotted lines) the two curves follow very similar trajectories. In both conditions, Americans cross the phoneme boundary at the same point, between stimuli 5 and 6. The differences are not significant, $\chi^2_{(1)} = 0.012$, $p > 0.05$.

As for identification of [ɑ] – [ɛ] in isolation and in context, Figures 3 and 4, a greater difference between the curves across both groups is seen. Resulting functions for the Catalan responses (the solid line) in the two conditions are,

ISBN 1-876346-48-5 © 2003 UAB

as predicted, significantly different, $\chi^2_{(1)} = 53.708$, $p < 0.01$. Catalans do not begin hearing front vowels until the very end of the continuum in context, between stimuli 7 and 8; while in isolation, front vowels begin to be heard at a more intermediate point in the curve. Stimulus 6 is associated with 50% front vowel identification. Again, we attribute the significant differences between the vowel in context and the vowel in isolation continua to the lack of experience with lip rounded alveolo-palatal fricatives in the native language. Vowels in context were not perceived in the same way as vowels in isolation. The vowels in context were identified as being significantly more 'back,' as a perceptual result of the lowering of F2 and F3. For Americans (represented by the dotted lines), the two identification functions are similar aside from the early stimuli in isolation. It seems that [ɑ] in isolation was more difficult to categorize than in context. Despite this, Americans do begin to identify front vowels at the same point in the two conditions, between stimuli 4 and 5. Because of the apparent lack of clarity of the vowel in isolation continuum however, disparity is great enough between the two conditions to reach significance at the 5% level, $\chi^2_{(1)} = 5.129$, $p < 0.05$. The trend though appears to be one towards insignificance, once the problematic stimuli early in the isolation continuum are discounted.

5. CONCLUSION

This experiment shows that cross-linguistic differences in coarticulatory effects are relevant in speech perception and provides further evidence that listeners are able to factor out coarticulatory variation if they have experience with such variation, as in the case of the American subjects. In the event of unfamiliar variation, however, listeners typically fail to normalize the effects induced by the unpredictable source instigating variation and give a face value interpretation to the signal, as witnessed by the Catalan subjects' performance.

As we have argued in this paper, Catalan speakers appear to have treated the low F2 and F3, caused by [ʃ̹], as being a property of the vowel rather than a consequence of lip rounded alveolo-palatal fricatives and thus identified fewer front vowels in context vis-à-vis American speakers. A further result of the Catalans' face value interpretation is that, as predicted, this particular group yields two differing identification functions for vowels in isolation and in context. Americans, on the other hand, who we argue were able to factor out the effects of lip rounding and identify the vowel similarly in context and in isolation, show insignificantly different trends for identification of vowels in context and in isolation.

ACKNOWLEDGEMENTS

Work supported by grant BFF2001-2498 from the Ministry of Education, Spain.

REFERENCES

[1] P.S. Beddor, R.A. Krakow, and L.M. Goldstein, "Perceptual constraints and phonological change: A study of nasal vowel height," *Phonological Yearbook* 3, pp. 197-217, 1986.

[2] R.A. Krakow, P.S. Beddor, L.M. Goldstein, and C. Fowler, "Coarticulatory influences on the perceived height of nasal vowels," *The Journal of the Acoustical Society of America* 83(3), pp. 1146-1158, 1988.

[3] O. Kunisaki and H. Fujisaki, "On the influence of context upon perception of fricative consonants," *Research Institute of Logopedics and Phoniatrics Annual Bulletin,* University of Tokyo, vol. 11, pp. 85-91, 1977.

[4] V.A. Mann and B.H. Repp, "Influence of vocalic context on perception of the [ʃ]-[s] distinction," *Perception and Psychophysics* 28, pp. 213-228, 1980.

[5] J.J. Ohala, "The listener as a source of sound change," in *Papers from the Para session on Language and Behavior,* C.S. Masek, R.A. Hendrick, and M.F. Miller, Eds., pp. 178-203. Chicago: Chicago Linguistic Society, 1981.

[6] D.B. Pisoni, S.E. Lively and J.S. Logan, "Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception," in *Development of Speech Perception: The transition from sounds to spoken words,* J. Goodman and H.C. Nusbaum, Eds., pp. 121-166, 1994.

[7] D. Recasens, *Fonètica Descriptiva Del Català*, Barcelona: Institut D'Estudis Catalans, 1991.

[8] C.D. Schatz, "The role of context in the perception of stops," *Journal of the Linguistic Society of America* 30, vol. I, 1954.

[9] M.J. Solé, "Phonetic & phonological processes: the case of nasalization," *Language and Speech* 35(1, 2), pp. 29-43, 1992.

[10] Y. Xu, "Production and perception of coarticulated tones," *The Journal of the Acoustical Society of America* 95(4), pp. 2240-2253, 1994.