ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2012
Sixth International Conference
Shanghai, China
May 22-25, 2012

Compensation for coarticulation in prosodically weak words

Felicitas Kleber, Jonathan Harrington, Ulrich Reubold, Jessica Siddins

Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University Munich, Germany
{kleber|jmh|reubold|jessica}@phonetik.uni-muenchen.de

Abstract

The goal of this study was to compare production differences in consonant-on-vowel coarticulation in accented and unaccented words with the extent to which listeners compensate differently for these coarticulatory effects. Native speakers of German produced nonsense words such as /pʊp/, /pʏp/, /tʊt/, and /tʏt/ that were either accented or unaccented. In a perception experiment, the same speakers made vowel quality judgements in /pʊp – pʏp/ and /tʊt – tʏt/ continua embedded in contexts in which they were prosodically strong or weak. Consonant-on-vowel coarticulation was found to be greater in the production of prosodically weak words. Listeners compensated for these coarticulatory effects in both prosodic conditions to approximately the same extent. These findings indicate that there was a mismatch between the production and perception of coarticulation in unaccented words. The results are discussed in terms of increased likelihood for sound change to occur in prosodically weak contexts.

Index Terms: compensation for coarticulation, accentuation, production-perception relationship, sound change

10.21437/SpeechProsody.2012-78

1. Introduction

Accented words are more likely to be hyperarticulated since very often they cannot be predicted from the context and signal new information [1]. On the other hand, unaccented words frequently carry old information and are predictable, and therefore tend to be hypoarticulated, as some of the acoustic information in the signal is semantically redundant. There is also some evidence to show that the magnitude of coarticulation is greater in hypoarticulated than in hyperarticulated productions [2].

According to Ohala [3], the great deal of synchronic variability does not lead to sound change because listeners are typically adept at compensating perceptually for the effects of coarticulation [4]. However, sound change may result in certain contexts in which the transmission between a speaker and hearer of coarticulatory information is inherently ambiguous [5]. Synchronically, for example, back vowels tend to be fronted to a central position in alveolar contexts because the tongue dorsum is advanced under the influence of the alveolar contact. However, if listeners fail to attribute this fronting to the coarticulatory source from which it originates (the alveolar consonants) then they might instead parse it with the vowel resulting potentially in the diachronic fronting of back vowels [6].

As far as the association between prosodic accentuation and sound change is concerned, there is extensive evidence to show firstly that sound change is frequent in prosodically weak contexts and secondly that this type of sound change can be related to increasing gestural overlap that occurs in production: for example, the synchronic basis of the historical derivation of present-day English *monks* from Old-English *muneceas* may be an increased gestural hiding of the weak vowel by the neighbouring consonants [7, 8].

A largely explored issue – that is also the main subject of this paper – is whether perceptual factors also contribute to such sound changes. More specifically, listeners might undercompensate for the high degree of coarticulation that occurs in prosodically weak contexts (see also [9]). Under this scenario, listeners might perceptually compensate less for coarticulation in prosodically weak as opposed to strong contexts, even though the magnitude of coarticulation in hypoarticulated, prosodically unaccented words is likely to be greater than in their accented counterparts. Sound change following Ohala's [3] model would then be more likely in the unaccented context precisely because the perceptual compensation for coarticulation would be too small in relation to the magnitude of coarticulatory influences in production. Thus the overall goal was to test whether a larger mismatch between the production and perception of coarticulation might explain the higher prevalence of sound change in prosodically weak contexts. This goal was formulated as the following three hypotheses:

(H1) There is more coarticulation in prosodically unaccented words than in accented words.

(H2) Listeners compensate perceptually for the effect of stop-to-vowel coarticulation.

(H3) Listeners compensate less for coarticulation in prosodically weak words.

2. Method

We tested these three hypotheses with respect to the synchronic fronting effects of alveolar place of articulation on high back vowels referred to earlier in production and perception.

2.1. Participants

Speech recordings and perception data were obtained from 10 speakers of Standard German. None of the subjects reported any hearing, eyesight or reading disabilities.

2.2. Materials

2.2.1. Production

The test words were symmetrical monosyllabic nonsense words with either /ʊ/ or /ʏ/ in the nucleus and flanked by /p/ or /t/ consonants. The words were embedded in the carrier phrase *Maria hat* _____ *gesagt* (*Mary said* _____ .) and presented to the participants together with 12 distracter sentences of an analogous structure in orthographic form. In order to elicit both unaccented and accented target words in production, subjects read sentences as answers to two different questions that were presented immediately before the test sentence. For the prosodically accented condition, the context was *Was hat Maria gesagt?* (*What did Mary say?*) which would elicit a pitch accent on the target word in the answer. For the unaccented condition, subjects saw the question *Wer hat T gesagt?* (*Who said T?*) where *T* is one of the four target words. We predicted that the answer to this question would be

produced such that the nuclear pitch accent fell on *Maria* with the following target word being deaccented. In addition, the word to be accented was printed in capital letters.

2.2.2. Perception

For the perception experiment, we created an 11-step continuum between natural realizations of /pʊp/ and /tʏt/ produced by a phonetically trained speaker using the static morphing method supplied by the AKUSTYK software add-on in *Praat* [10]. The resulting F2 values for each stimulus are shown in Table 1. We then spliced the 11 vowels into two contexts: labial /p_p/ and alveolar /t_t/. For the prosodically accented condition, each of these CVC stimuli was embedded in the carrier sentence *Maria hat CVC gesagt* with one H* pitch accent on CVC (the final boundary tones were L-L%). For the unaccented condition, only *Maria* was (nuclear) accented and the CVC was post-focal and deaccented. The sentence in the unaccented condition was derived from that of the accented condition by shifting the f0 peak to Maria and flattening the f0 contour on the target word using the manipulation and overlap-add resynthesis function in Praat; in addition, we lengthened the final two syllables of *Maria* (/ri:/ carries the primary stress), and raised and lowered the intensity by 5 dB over the intervals of *Maria* and the CVC target word, respectively.

Table 1. *F2 [Hz] values for each stimulus number.*

Stimulus	F2 [Hz]
1	803
2	808
3	861
4	956
5	989
6	1088
7	1121
8	1239
9	1310
10	1328
11	1436

2.3. Experimental set-up

The recording took place before the perception experiment, in order not to draw subjects' attention to the target words, which were obscured by filler words in the production task only. Both the production and perception data were obtained in one session per speaker in a sound attenuated booth at the Institute of Phonetics and Speech Processing, Munich.

Ten repetitions of each target and filler sentence in both prosodic conditions, i.e. with both questions, were presented in randomized order on a computer screen. Subjects were instructed to carefully and silently read the question and then to read aloud the answer with the accentuation on the corresponding and highlighted word. In case of false pronunciations and/or accentuation patterns, the subject was asked to read the answer again. There was a total of 200 sentences ((four test words + six filler words) x two accentuation patterns x ten repetitions) and the recording session lasted approximately 30 minutes. The subjects were free to take a break whenever they needed one.

The perception experiment was conducted using *Praat*. Each of the 44 stimuli (11 F2-steps x two consonantal contexts x two accentuation patterns) was repeated ten times and presented in randomized order to the subjects via headphones. For each auditory stimulus, the subject saw two corresponding word alternatives in orthographic form differing only in the nucleus, i.e. <u> or <ü> (corresponding to /ʊ/ and /ʏ/, respectively), on a computer screen; the presentation order on the screen was counterbalanced. The subjects' task was to click on the word they had perceived. The next stimulus was presented with a delay of one second after the previous response.

2.4. Data analysis

The data was manually segmented in *Praat*. The vowel's onset and offset were defined by the beginning and end of F2. The beginning of the preceding stop (C1) was either set after the burst of the /t/ in *hat* if it was released or at the temporal midpoint of the closure phase if there was no visible /t/-release in *hat.* The offset of the target word's final stop (C2) was placed at the end of the aspiration. Each utterance was checked for a correct accentuation pattern, i.e. one pitch accent on the target word and deaccentuation of the name or vice versa. The production data of one subject had to be excluded because she produced only tense vowels. All data files were then converted into EMU Speech Database System files and all further analyses were carried out in EMU/R [11]. The first four formants were calculated in EMU with the following parameters: LPC order of 10, a pre-emphasis of 0.95, and a 30 ms Blackman window with a frame shift of 5 ms. The formant data was checked manually and corrected if necessary.

Word, vowel and stop durations were extracted in EMU/R and vowel duration was normalized for word duration. The first and second formant were measured at the vowel's temporal midpoint and converted to Bark using the formula in [12]. F0 was extracted in EMU and the median for each vowel was calculated in EMU/R. The production data was then analyzed using general linear mixed models (GLMM) with either word or vowel duration, f0, and either F1 or F2 as the dependent variable and Accentuation (two levels: accented vs. unaccented), Vowel (two levels: /ʊ/ vs. /ʏ/), and Place of articulation (two levels: alveolar vs. labial) as fixed factors and Speaker as a random factor.

The perception data of four subjects (including the subject that was excluded from the production analysis) were excluded since they showed no categorical shift in their perception of the labial continuum. The perception data was analyzed by means of a GLMM which is described in detail below in 3.2.

3. Results

3.1. Speech recordings

F0 was higher in accented words than in unaccented words (cf. Figure 1). A GLMM with the median f0 as the dependent variable revealed a significant main effect for Accentuation ($F[1,60]=640.5$, $p < 0.001$) and no other significant effects.

Word duration was significantly longer in accented than in unaccented words. This was mainly due to a lengthening of the preceding and following plosive, in particular when the stop's place of articulation was alveolar, and only partly due to longer vowel durations in accented words (cf. Figure 2). A GLMM with absolute word duration as the dependent variable revealed significant main effects for Accentuation

(F[1,60]=184.1, p < 0.001) and Place of articulation (F[1,60]=25.1, p < 0.001). This result indicates that word duration differs significantly depending on the stop's place of articulation and whether or not the word is accented.

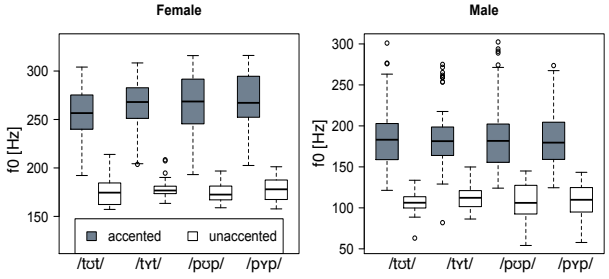

Figure 1: *Median f0 [Hz] in vowels of accented (grey) and unaccented (white) words across all speakers, separately for vowel and place of articulation.*

A second GLMM with proportional vowel duration as the dependent variable showed a significant main effect for Place of articulation (F[1,60]=26.1, p < 0.001) and a significant interaction between Vowel x Place of articulation (F[1,60]=24.9, p < 0.001), indicating that proportional vowel duration was significantly greater in alveolar than in labial contexts, in particular in tokens that contained the vowel /ʊ/. However, it was not affected by Accentuation.

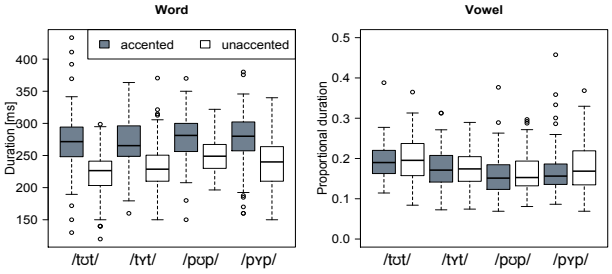

Figure 2: *Absolute word duration and proportional vowel duration of accented (grey) and unaccented (white) words across all speakers, separately for vowel and place of articulation.*

In all four contexts, F1 was lower in unaccented words than in accented words (cf. Figure 3). A GLMM with F1 at the vowel's temporal midpoint as the dependent variable revealed significant main effects for Vowel (F[1,60]=35.5, p < 0.001), Accentuation (F[1,60]=213.7, p < 0.001), and Place of articulation (F[1,60]=7.1, p < 0.01) as well as a significant interaction between Vowel x Place of articulation (F[1,60]=21.2, p < 0.001). The significant effect for Accentuation points to a decreased degree of jaw opening (F1 lowering) when the words were deaccented.

There was no or only little target undershoot in unaccented words (cf. Figure 3) except for /tʊt/ in male speakers, where a higher F2 target was reached in unaccented position which indicates that the consonant-on-vowel coarticulation in /tʊt/ was greater in the prosodically weak condition. A GLMM with F2 at the midpoint of the vowel as the dependent variable showed significant main effects for Vowel (F[1,60]=9139.6, p < 0.001), Accentuation (F[1,60]=11.5, p < 0.01), and Place of articulation (F[1,60]=1773.0, p < 0.001) as well as significant interaction effects between Vowel x Accentuation (F[1,60]=7.1, p < 0.01), Vowel x Place of articulation (F[1,60]=275.3, p < 0.001), and Accentuation x Place of articulation (F[1,60]=9.5, p < 0.01). The latter two interactions indicate that the degree of consonant-on-vowel coarticulation was significantly greater for prosodically weak words as opposed to strong words.

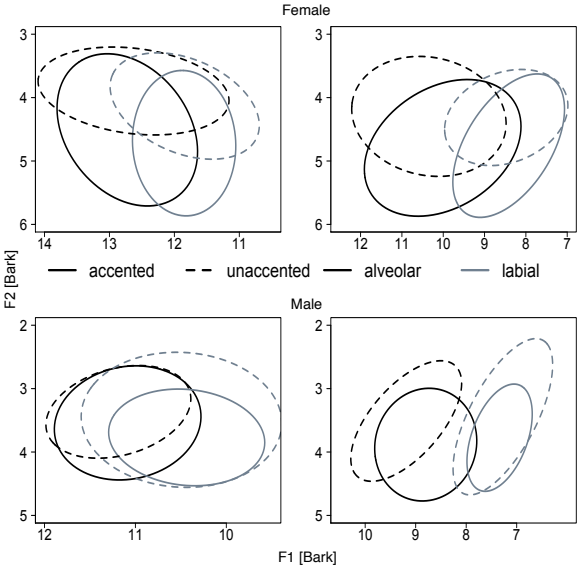

Figure 3: *95% confidence ellipses on a Bark-scaled F2 x F1 plane for /ʏ/ (left) and /ʊ/ (right) in alveolar (black) and labial (grey) as well as accented (solid) and unaccented (dashed) contexts, shown separately for female (top) and male speakers (bottom).*

To summarize the results, speakers used f0 and duration in order to mark accentuation. F1 was affected mainly by Accentuation and F2 primarily by Place of articulation due to the consonant-on-vowel coarticulation.

3.2. Forced choice identification test

Figure 4 shows the psychometric response curves to the four continua fitted to the response data of 6 subjects using a GLMM with Stimulus response as the dependent variable and Accentuation (two levels: accented vs. unaccented), Vowel (two levels: /ʊ/ vs. /ʏ/), and Place of articulation (two levels: alveolar vs. labial) as fixed factors and Listener as random factor. The result of this operation was to fit a logistic function to the stimulus responses (separately by listener) using the relationship

$$p = \frac{e^{(mx+k)}}{1+e^{(mx+k)}} \quad (1)$$

where p was the predicted proportion of /ʏ/-responses (0 < p < 1), the coefficients m (the *slope*) and k (the *intercept*) were calculated separately for each continuum and listener, and x was the stimulus number 1,2, …11.

A GLMM with the 50% cross-over boundary, calculated by *-m/k*, as the dependent variable and the same fixed and random effects as above revealed a significant main effect for Place of articulation (F[1,60] = 193.2, p < 0.001) and no other significant effects. Commensurate with Figure 4, listeners perceived significantly more stimuli from the /ʊ – ʏ/

continuum as /ʏ/ when the vowel occurred in the labial context as opposed to the alveolar context, i.e. the phoneme boundary was left-shifted in /pʊp – pʏp/ compared to /tʊt – tʏt/. This means that listeners compensated for coarticulation by attributing higher F2 values to the consonantal context in /tʊt – tʏt/ but not in /pʊp – pʏp/. On the other hand, whether or not the target word was accented did not influence the placement of the phoneme category boundary.

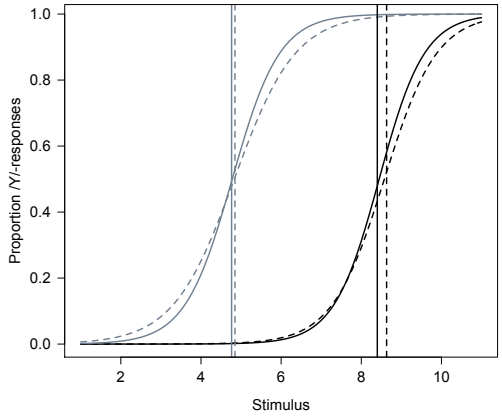

Figure 4: *Proportional distribution of* /ʏ/*-responses in alveolar (black) and labial (grey) contexts in accented (solid) and unaccented (dashed) conditions. The vertical lines indicate the mean category boundaries between* /ʊ/ *and* /ʏ/.

4. Discussion & Conclusions

Speakers produced accented words with longer durations, a higher f0 and a higher F1. These findings replicate earlier results showing longer segment durations, higher f0 and higher F1 values in prosodically strong words [13, 14]. Second, the effects of consonant-on-vowel-coarticulation, such as /ʊ/-fronting in an alveolar environment which results acoustically in higher F2 values, were greater in unaccented than accented words. This finding is commensurate with hypothesis H1. Third, listeners compensated for the effect of coarticulation since they interpreted some of the F2 raising as context-induced when the stop's place of articulation is alveolar. This result confirms the second hypothesis H2 and is consistent with the findings of a number of earlier studies showing perceptual compensation for coarticulation [4, 5].

Our third and most important hypothesis H3 was that listeners compensate less for coarticulation in prosodically weak words. Our findings do not support this hypothesis. Listeners compensated for coarticulatory effects to the same extent in unaccented words as they did in accented tokens. However – and this is commensurate with a hypothesis that predicts a greater mismatch between perception and production of coarticulation in unaccented words – listeners did *not* compensate *more* for coarticulation in prosodically weak contexts although their production showed increased coarticulation in unaccented words.

Unaccented words are likely to be produced with a greater amount of coarticulation since they tend to be hypoarticulated [1]. At the same time, however, listeners are less able to pay attention to the fine phonetic detail of hypoarticulated words in prosodically weak positions. Therefore, the probability with which listeners fail to correctly parse coarticulation as intended by a speaker increases in unaccented words. According to Ohala, sound change is driven by the misperception of coarticulation [3]. Consequently, sound changes should occur more often in prosodically weak contexts, which are prone to more coarticulation. In this study we have presented some evidence that the mismatch between production and perception of coarticulation is magnified in prosodically weak words. This mismatch in unaccented words, then, might be the reason for incorrect parsing of coarticulatiory patterns which can result in sound changes.

5. References

[1] B. Lindblom, "Explaining phonetic variation: a sketch of the H&H theory," in *Speech production and speech modelling*, W. J. Hardcastle and A. Marchal, Eds., 1990, pp. 403-439.

[2] C. A. Fowler, "Parsing coarticulated speech in perception: effects of coarticulation resistance," *Journal of Phonetics,* vol. 33, pp. 199-213, 2005.

[3] J. J. Ohala, "Sound change as nature's speech perception experiment," *Speech Communication,* vol. 13, pp. 155-161, 1993.

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

[5] P. S. Beddor, A. Brasher, and C. Narayan, "Applying perceptual methods to the study of phonetic variation and sound change," in *Experimental Approaches to Phonology*, M. J. Solé, P. S. Beddor, and M. Ohala, Eds., Oxford: Oxford University Press, 2007, pp. 127–143.

[6] J. Harrington, F. Kleber, and U. Reubold, "Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: an acoustic and perceptual study," *Journal of the Acoustical Society of America,* vol. 123, pp. 2825-2835, 2008.

[7] M. E. Beckman, K. de Jong, S.-A. Jun, and S.-H. Lee, "The Interaction of Coarticulation and Prosody in Sound Change," *Language and Speech,* vol. 35, pp. 45 - 58, 1992.

[8] C. P. Browman and L. M. Goldstein, "Gestural structures: distinctiveness, phonological processes, and historical change.," in *Modularity and the Motor Theory of Speech Perception*, I. Mattingly and M. Studdert-Kennedy, Eds., New Jersey: Erlbaum, 1991, pp. 313 - 338.

[9] B. Lindblom, S. Guion, S. Hura, S.-J. Moon, and R. Willerman, "Is sound change adaptive?," *Rivista Di Linguistica,* vol. 7, pp. 5-37, 1995.

[10] P. Boersma and D. Weenink, "Praat: doing phonetcs by computer.," 5.2.21, 2011.

[11] J. Harrington, *The Phonetic Analysis of Speech Corpora*. Chichester: Wiley-Blackwell, 2010.

[12] H. Traunmüller, "Analytical expressions for the tonotopic sensory scale," *Journal of the Acoustical Society of America,* vol. 88, pp. 97 - 100, 1990.

[13] T. Cho, "Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English," *The Journal of the Acoustical Society of America,* vol. 117, pp. 3867 - 3878, 2005.

[14] I. Lehiste, *Suprasegmentals*. Cambridge, Mass.,: M.I.T. Press, 1970.