THE INFLUENCE OF SLOWED SPEECH RATE ON COARTICULATION: ACOUSTIC ANALYSIS OF DURATIONAL AND SPECTRAL PARAMETERS

Ingo Hertrich and Hermann Ackermann
University of Tübingen, Germany

ABSTRACT

Durational and spectral measures of coarticulation were obtained from six young female subjects. Anticipatory and retentive coarticulation differed in their acoustic patterns. Both effects showed considerable inter-subject variability. Slowing of speaking rate resulted in a decrease of retentive coarticulation only. The data corroborate the suggestion of different mechanisms underlying anticipatory and retentive coarticulation.

INTRODUCTION

During running speech neighbouring phonetic segments affect each other in various ways, which are usually referred to as coarticulation [1] [2]. Anticipatory as well as retentive coarticulation have been observed [2] [3] [4] [5]. There is some evidence that accelerated speech gives rise to increased coarticulation [6] [7] [8]. This influence of fast speaking rate on coarticulation has been explained by an increased overlap of the temporal domains or activation fields of adjacent articulatory gestures [9] [10] [11]. In analogy, reduced coarticulatory effects may be expected in slow speech.

In order to assess the influence of slowed speech rate upon anticipatory and retentive coarticulation, the acoustic signal of German sentence utterances was analyzed with respect to durational and spectral parameters. Most available acoustic studies on coarticulation considered the formants of the speech signal. However, formant analysis may be compromised by inherent shortcomings. First, it relies on the assumption that the frequency values of a limited number of distinct spectral peaks convey the relevant information. Second, formant extraction algorithms may split a single formant into two or combine two formants to a single one giving rise to incorrect assignments. In order to avoid these problems the present study relied on averaged FFT spectra without any correction or decision making algorithms such as formant tracking.

METHODS

Subjects and material

Six young females participated in the present study representing a rather homogeneous group with respect to age, sex, and education.

Three test sentences comprising the nonsense target word "getVte" (V = {a | i | u}) embedded into the German carrier phrase "Ich habe .... gelesen" ("I have read ....") were considered for analysis. The vowel /V/ of the target word "getVte" has the most prominent position. Thus coarticulation effects of /V/ on both the preceding and the succeeding /e/ and /t/ segments may be expected.

Procedure

Each of the three target words was printed on a card in bold letters. The subjects had to produce the respective test sentences ten times each at comfortable speech tempo (normal speech rate), five times with the instruction to speak somewhat slower (slow-condition) and five times at an even slower rate (extra slow-condition).

Acoustic processing

All recorded sencentes were digitized at a sampling rate of 20 kHz (after anti-aliasing filtering at 8 kHz). Analysis was performed with the CSL Speech Lab (CSL4300; KAY Elemetrics, USA).

Vowel onsets, vowel ends, and stop consonant bursts of the produced test sentences were marked (Figure 1). The following segment durations were considered for analysis: pre-accent /e/ (=SCHWA1), occlusion of pre-accent /t/ (=OCC1), voice onset time of pre-accent /t/ (=VOT1), target vowel (=V) occlusion of post-target /t/ (=OCC2), voice onset time of post-target /t/ (=VOT2), and post-target /e/ (=SCHWA2).

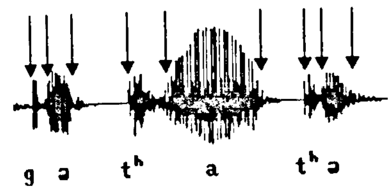

Figure 1. Target word "getate" (Vowel on- and offsets as well as bursts are marked by arrows)

Spectral analysis was performed with the following five segments (bold letters in "Ich habe **g-e-t-V-t-e** gelesen"): pre-target /e/ (=SCHWA1), aspiration of pre-target /t/ (=ASP1), target vowel (=V), aspiration of post-target /t/ (=ASP2), and post-target /e/ (=SCHWA2).

From each of the five segments an averaged FFT spectrum was made (frame length = 6.4 ms; frequency resolution = 156.25 Hz). Altogether 6 (subjects) x 5 (segments per sentence) x 3 (target vowels) x 20 (repetitions: 10 with normal rate, 5 slow, 5 extra slow) = 1800 averaged spectra were computed. The spectra, given in dB values on a linear frequency scale, were normalized for overall intensity by subtracting the mean dB value from all frequency bands.

Statistical analysis

SAS 6.03 software was used for statistical analysis (SAS Institute Inc.; Cary/NC, USA). Multivariate analyses were performed in order to test the differential contribution of the three factors SUBJ {N1-N6}, RATE {normal | slow | extra slow}, and TARG {a | i | u} to the variability of the durational and spectral parameters. The seven segment durations (SCHWA1, OCC1, VOT1, V, OCC2, VOT2, SCHWA2) were converted onto a logarithmic scale in order to compensate for the increase of variances with absolute segment durations [12].

With the spectral data a principal component analysis was performed as a first step in order to reduce the large number of variables. Post hoc analyses of spectral energy distribution were performed in order to find out the frequency bands relevant for the observed multivariate effects. To these ends spectra were averaged across repetitions and plotted together with the corresponding two-standard-error bars. This procedure allowed to identify the spectral regions of interest with respect to coarticulatory patterns.

RESULTS

Segment durations

Seven multivariate tests (the three main effects of TARG, RATE, and SUBJ and the four possible interactions) were performed with the segment durations as the numeric variables. All three factors showed significant effects. As expected, the rate condition revealed to be the dominant source of variation. Subject variability exceeded the influence of coarticulation effects. The two-factor interaction RATE x TARG was not significant when the duration of the target vowel itself was excluded from the numerical data set, indicating that the durational coarticulatory patterns, did not vary across speaking rate conditions.

The RATE factor revealed to be the dominant source of variation with respect to all segments except VOT1, the latter representing a primarily subject-specific measure. The factor TARG had its largest influence upon /V/ reflecting the intrinsic target duration: /a/ had the longest duration, /i/ was slightly shorter than /u/. Significant coarticulatory influences of TARG were observed upon OCC1, VOT1, and SCHWA2. In all subjects the relatively long intrinsic duration of the target vowel /a/ was in part compensated by reduction of one or more of the adjacent segments, predominantly the pre-accent occlusion (OCC1).

In summary, the durational analyses yielded the following results: (1) The three target vowels showed intrinsic durational variability. (2) Partially the coarticulation effects compensated the intrinsic target vowel durations. (3) The durational coarticulation patterns did not significantly interact with the speech rate condition.

Spectral analysis

For reasons of data reduction a principal component analysis was performed with all spectra up to 6 kHz (= 39 spectral bands). Components nos. 1 to 8, accounting for 88% of total variance, were used as dependent numerical variables for multivariate analyses. The categorical variables SUBJ, TARG, and RATE represented the independent factors. A separate MANOVA was performed with each of the the five speech segments SCHWA1, ASP1, V, ASP2, and SCHWA2.

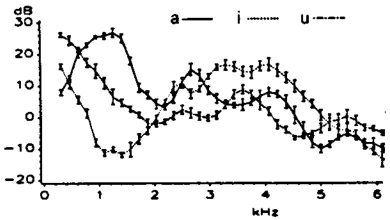

Figure 2. Target vowels /a/, /i/, /u/, (means with 2-standard-error bars across 10 repetitions)

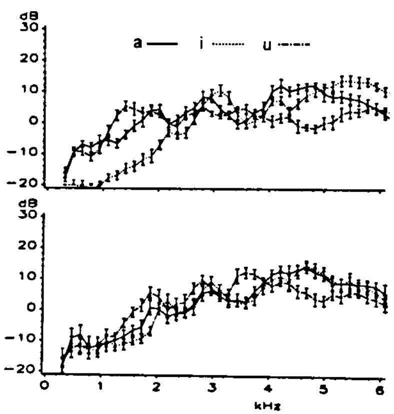

Figure 3. Coarticulation effect upon the pre-target (top) and the post-target (bottom) aspiration (means with 2-standard-error bars across 10 repetitions)

Apart from a few interactions all comparisons yielded significant results. As exemplified in Figure 2, the most significant effect was the (trivial) one of the factor TARG upon the target vowel itself.

The influence of TARG upon the pre-target aspiration had almost the same strength and exceeded the TARG-effect upon ASP2 (Figure 3). Whereas the target vowel category was the dominant source of variation of the aspiration segments, subject variability revealed to be the prevailing effect with respect to the two schwa-sounds. With respect to SCHWA1 some subjects did not show any coarticulatory behavior even under the normal speech rate condition (Figure 4).

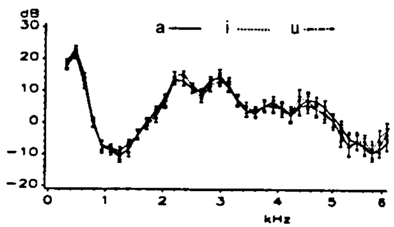

Figure 4. Absence of any coarticulation effect of TARG upon SCHWA1 (means with 2-standard-error bars across 10 repetitions of one subject under the normal rate condition)

The interactions TARG x RATE and TARG x RATE x SUBJ failed significance with respect to SCHWA1. Thus, anticipatory vowel-to-vowel coarticulation did not systematically depend on the speech rate condition. In contrast, SCHWA2 showed reduced coarticulation under the slow rate conditions in all subjects as exemplified in Figure 5.

With respect to all segments considered the interaction between target vowel category and speech rate condition (TARG x RATE) was weaker than the interaction between target vowel category and subject variability (TARG x SUBJ). This indicates that the coarticulatory patterns depend stronger on individual factors than on the speech rate condition.

To summarize the results of the spectral analyses:

(1) The pre-target consonant aspiration (ASP1) showed the strongest coarticulation effect, followed - in descending order - by post-target aspiration (ASP2), post-target schwa (SCHWA2), and pre-target schwa (SCHWA1).

(2) The target vowel category was the dominant source of the observed variability of the aspiration spectra. In contrast, the schwa-sounds primarily were characterized by individual patterns.

(3) Slowing of speech tempo had only minor influences on the aspiration segments. The schwa-segments, in contrast, showed stronger rate effects.

(4) Anticipatory vowel-to-vowel coarticulation did not systematically interact with the speech rate condition.

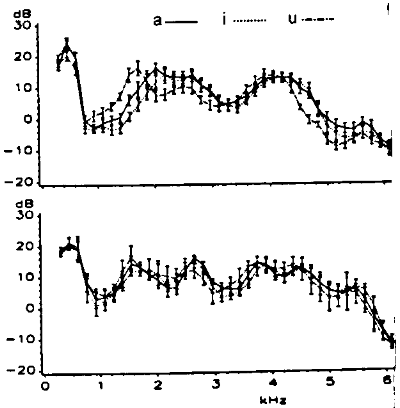

Figure 5. Coarticulation effect upon the post-target /e/ (SCHWA2) under the normal (top) and the extra slow (bottom) speech rate condition (means with 2-standard-error bars across repetitions)

CONCLUSIONS

Durational coarticulatory influences seem to reflect, at least in part, a compensation for intrinsic differences in target length. Spectral analysis showed that retentive coarticulation was weakened in slow speech, whereas anticipatory coarticulation was not. The presence of anticipatory vowel-to-vowel coarticulation was speaker-specific.

REFERENCES

[1] Daniloff, R.G., and Hammarberg, R.E. (1973), On defining coarticulation. *Journal of Phonetics*, vol. 1, pp. 239-248.

[2] Kent, R.D., and Minifie, F.D. (1977), Coarticulation in recent speech production models. *Journal of Phonetics*, vol. 5, pp. 115-133.

[3] Kent, R.D., and Moll, K.L. (1972), Tongue body articulation during vowel and diphthong gestures. *Folia Phoniatrica*, vol. 24, pp. 278-300.

[4] Parush, A., and Ostry, D.J. (1993), Lower pharyngeal wall coarticulation in VCV syllables. *Journal of the Acoustical Society of America*, vol. 94, pp. 715-722.

[5] Repp, B.H. (1986), Some observations on the development of coarticulation. *Journal of the Acoustical Society of America*, vol. 79, pp. 1616-1619.

[6] Gay, T. (1968), Effect of speaking rate on diphthong formant movements. *Journal of the Acoustical Society of America*, vol. 44, pp. 1570-1573.

[7] Gay, T. (1974), Effect of speaking rate on stop consonant-vowel articulation. *Paper presented at the Speech Communication Seminar, Stockholm, Aug. 1-3, 1974.*

[8] Gay, T., Ushijima, T., Hirose, H., and Cooper, F.S. (1974), Effect of speaking rate on labial consonant-vowel articulation. *Journal of Phonetics*, vol. 2, pp. 47-63.

[9] Bell-Berti, F., and Harris, K.S. (1981), A temporal model of speech production. *Phonetica*, vol. 38, pp. 9-20.

[10] Fowler, C.A., and Saltzman, E. (1993), Coordination and coarticulation in speech production. *Language and Speech*, vol. 36, pp. 171-195.

[11] Löfqvist, A., and Yoshioka, H. (1981), Interarticulator programming in obstruent production. *Phonetica*, vol. 38, pp. 21-34.

[12] Crystal, T.H., and House, A.S. (1988), A note on the variability of timing control. *Journal of Speech and Hearing Research*, vol. 31, pp. 497-502.