SPEECH RATE EFFECTS ON DURATION: AN ARTICULATORY ANALYSIS

Corinne Roberts

Speech, Hearing, and Language Research Centre
Macquarie University

ABSTRACT - Duration differences resulting from speech rate changes are assessed from three perspectives: linear rescaling, gestural overlap and planned shortening. A method of analysis is provided to distinguish between linear rescaling and gestural overlap. The results of the research suggest a range of processes create the durational differences.

INTRODUCTION

The literature on articulatory variation across different speech rates provides a range of opinions on the processes underlying syllable lengthening and shortening. As was suggested by Gay (1981), it is perhaps intuitive to think of duration reduction as the compression of an utterance. Alternatively, the recent task dynamic model (Browman & Goldstein, 1990; Saltzman, 1991) suggests two main processes for the reduction of a syllable based on the dynamics of articulatory movements: interarticulatory overlap of neighbouring gestures and the intra-articulatory stiffness.

Lindblom (1963) associated smaller amplitudes and stable peaks in velocity at faster speech rates with target undershoot, and suggested that the duration reduction occurred as a result of physiological constraints. Wieneke, Janssen & Belderbos (1987) also found that jaw movements did not reach the same peaks in displacement at fast speeds as were found at slower rates. Whilst these results can be associated with the compression of speech, it is equally possible that adjacent gestures are overlapped. Intergestural overlap, or the truncation of the gestures, creates similar kinematic characteristics as for linear rescaling of the gestures (Harrington, Fletcher & Roberts, submitted). In order to assess these two theories the first section of this paper provides a method of analysing the differences between the linear rescaling of an utterance and the overlapping of the gestures.

An alternate view in the literature is that increasing speech rate does not produce a decrease in the displacement of the movement, but rather an increase in the peak velocity (Kuehn & Moll, 1976; Edwards, Beckman & Fletcher, 1991). It has been suggested that the increase in velocity results from an increase in the force with which the movement is produced, and that this increase in force reflects the "planned shortening" of the utterance from an increase in muscular stiffness (De Jong, 1991). The second section of the paper assesses the extent of planned shortening in reducing the duration.

METHOD

Three adult speakers of Australian English (female speakers: CR, KC, JF) were asked to produce the sentence "say *barb* naturally, not bub" at both a normal and fast rate. Fifteen repetitions of the nuclear accented syllable "barb" were elicited. The speech data from two of CR's tokens, and one of KC's, were not correctly aligned with the movement data, so these tokens are not included.

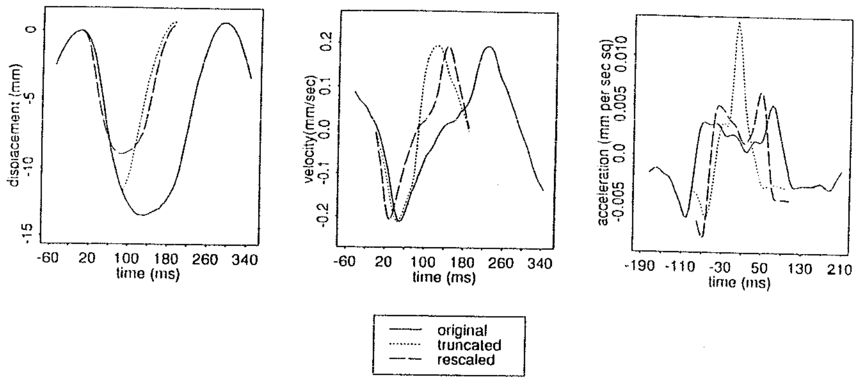

Figure 1. Trajectory differences from linear rescaling and gestural overlap.

The movements of the upper lip, lower lip and jaw were recorded using a Movetrack system with a sampling rate of 200 Hz, and the speech signal was recorded simultaneously with the acoustic sampling rate at 20 kHz. The phonetic boundaries were marked using Waves+ and subsequent analyses were executed using mu+. The opening gesture was determined as the movement from the minimum displacement of the initial /b/ to the peak displacement in the vowel, and the closing gesture from the peak displacement to the minimum displacement for the final /b/. Only jaw movements are analysed in this paper.

THE ANALYSIS OF LINEAR RESCALING COMPARED WITH GESTURAL OVERLAP

Although for rescaled and overlapped gestures the peak displacements and velocities are similar, it is possible to define three parameters on which the trajectories differ. Primarily the shape of the waveform between the peak velocities converges to a sharper peak in the truncated gesture than in the rescaled gesture (Harrington, Fletcher & Roberts, submitted). Figure 1 provides an example of the differences in the shape of the duration, velocity and acceleration as a result of artificial linear rescaling and gestural overlap (both duration and velocity are aligned at the time of the first minimum in displacement, and the acceleration is aligned at the time of the peak in displacement).

The first parameter looks at the duration between the peak velocities as a proportion of the syllable. As gestures in the syllable overlap more, the proportion of the syllables between the peak velocities decreases. This proportion remains the same when the syllable is rescaled.

The second metric is based on predicting the syllable duration from the ratio of the displacement to the peak velocity (Beckman, Edwards & Fletcher, 1992), assuming that the movement approximates a constant velocity. When the velocity is constant the time taken to complete the movement equals the peak displacement/velocity ratio. If the velocity deviates from the peak velocity, i.e. the movement slows or speeds up, a discrepancy is created between the predicted duration and the actual duration. This is applied to the difference between rescaled and truncated movements: the sharper the peak of the opening and closing gestures the more constant the velocity is, and consequently, the better the displacement/velocity ratio is as a predictor of the time it takes to complete the movement (refer to Figure 1). Thus it is expected that this ratio will be closer to the duration of the truncated syllables than the rescaled syllables. This parameter is calculated as the percentage of the syllable underestimated.

The final parameter considers the shape of the acceleration curve, which should show a peak corresponding to the peak displacement in the truncated gestures, where there would be no equivalent peak in the rescaled gestures. This results from the faster transition from the opening gesture to the closing gesture in the truncated syllable (see Figure 1). This parameter is quantified by taking the biggest peak in the acceleration curve, and determining the distance between this peak and the place of the peak displacement.

In order to determine whether the tokens produced at the fast rate are more like truncated or rescaled gestures, the normal syllables are truncated and rescaled by the average duration difference between the rates (speaker CR by 45 msec, KC by 90 msec, and JF by 55 msec), and compared with the fast tokens.

Results and discussion

Figure 2 shows a plot of the proportion of the syllable between the peak velocities and the proportion of the syllable which is underestimated. This figure shows that for speakers CR and KC

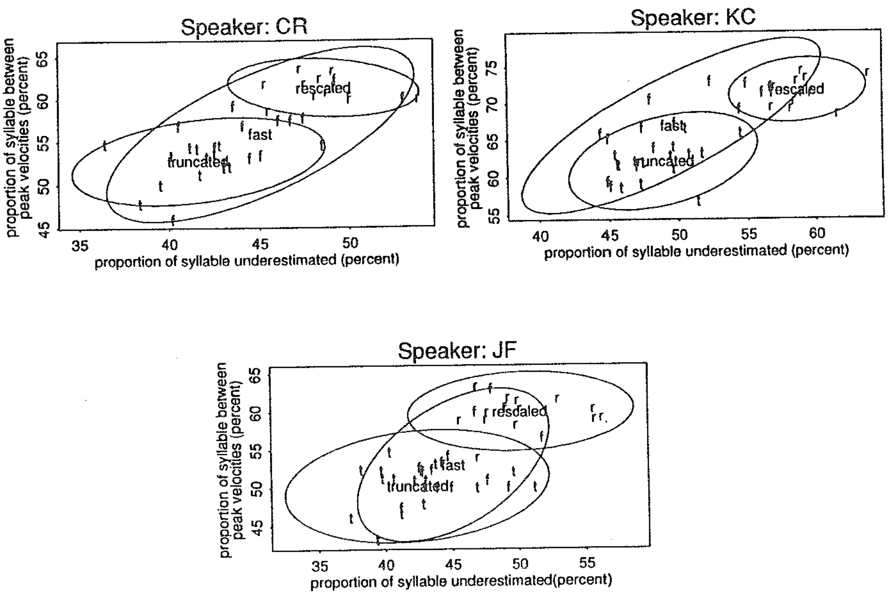

Figure 2. Ellipse plots including 95% of the fast, rescaled and truncated tokens.

the tokens produced at a fast rate conform to neither the truncated group nor the rescaled group. JF, however, is best modelled by the group of truncated tokens (the Bayesian distances from the fast tokens to the truncated centroid were less than the distances to the rescaled centroid; $t=2.27$, $p<0.05$). The acceleration traces in Figure 3 indicate that the fast tokens are more similar in shape to the rescaled group than the truncated group (the Bayesian distances from the fast tokens to the truncated mean, as opposed to the rescaled average distance, were not significant).

THE ANALYSIS OF PLANNED SHORTENING

Two characteristics of the movement trace are appropriate for assessing the extent of the planned shortening of fast syllables: the peak displacements and the peak velocities. If the speaker plans to decrease the duration of a syllable the result might be that the speaker speeds up the gestures while preserving the amplitude of the movement (De Jong, 1981). Thus the peaks in displacement would remain the same for different speech rates, whereas the peaks in velocity would increase as a result of producing the gestures with more force.

As the increase in muscular stiffness may differ between gestures, the opening and closing gestures are considered separately for each speech rate. The differences in duration for each gesture and rate are illustrated in Figure 4 so that any gestural effect can be considered. The peak displacement is measured at the point where the opening and closing gestures meet, so only one

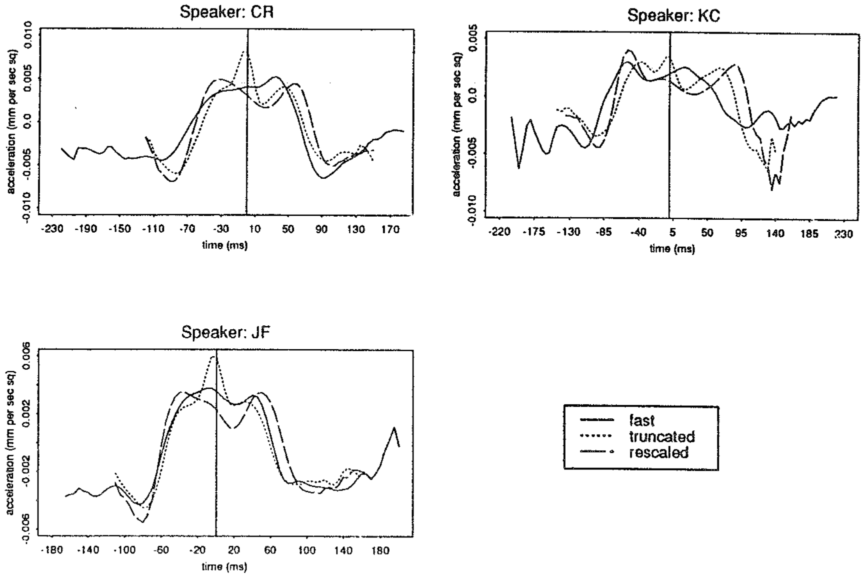

Figure 3. Averaged acceleration traces for fast, truncated, and rescaled tokens. The vertical line corresponds to $t=0$, the time of the peak displacement.

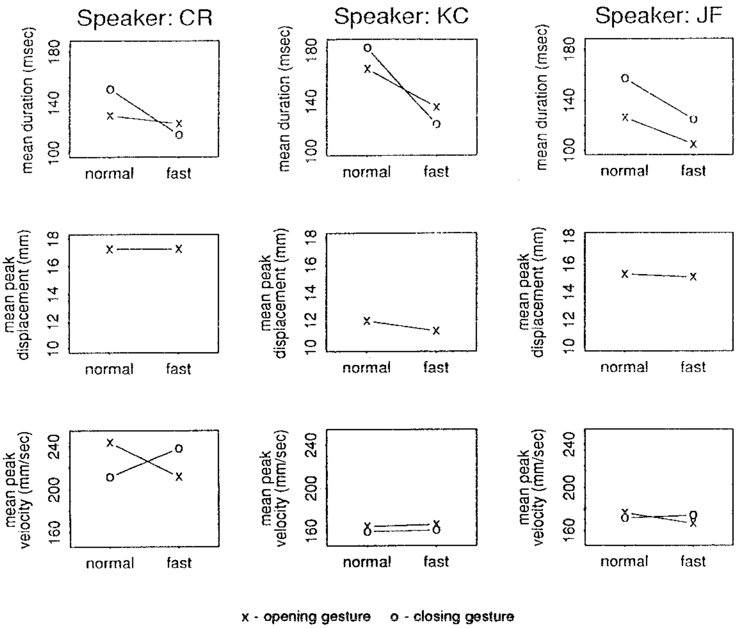

Figure 4. Mean durations, peak displacements, and peak velocities.

value is shown in Figure 4. The peak velocities, also shown in Figure 4, will be compared to determine any increase in force.

Results and discussion

The results of an analysis of variance (rate x gesture) show support for the planned shortening hypothesis as the peak displacements do not decrease significantly at the fast speech rate for any of the subjects (see Figure 4). Contrastingly, the results from a second analysis show the peak velocities neither increase nor decrease significantly as a result of the increase in rate, which indicates the movements may not be produced with an increase in force. The results from the rate by gesture interaction from speaker CR show a significant difference in the velocity dependent on which gesture of the syllable is involved (F=13.03, p<0.05), but no difference when considering the overall rate. While Figure 4 shows all subjects decrease duration, the closing gesture decreases more than the opening gesture (this interaction was significant for speaker CR, F=12.02, p<0.01; and for KC, F= 7.05, p<0.05), which is inconsistent with the patterns in velocity across the gestures.

CONCLUSION

Variation across subjects for the differences in rate is a commonly reported phenomenon (e.g. Kuehn & Moll, 1976; Gay, 1981), and it does not seem inconceivable that different people employ different strategies to reduce duration. From the support for the gestural overlap hypothesis shown by Speaker JF a wider study using the methodology presented for distinguishing between rescaling and truncation may provide clearer results.

The results from the second section of the paper are of some concern. If speakers are reducing the duration of the utterance, but are not decreasing the displacement in the reduction process, and additionally, they are not increasing the velocity in creating the duration difference, then what is changing to result in the reduction in duration?

The results suggest that the closing gesture's duration may decrease more than the opening gesture. The transition from the final closure into the next syllable could be having an effect on the closing gesture's duration. If, rather than raising the jaw fully for the final /b/, the speaker leaves the jaw quite low for the next gesture, then there would not necessarily be a decrease in the peak displacement of the /ba:b/ syllable, and nor would the velocity necessarily change. The closure for the final /b/ may be created by compensation from the lips, which has not been considered here. Further work on this topic could consider the intragestural coordination of the articulators at the point of closure in the final consonant.

REFERENCES

Beckman, M.E., Edwards, J. & Fletcher, J. (1992) *Prosodic structure and tempo in a sonority model of articulatory dynamics.* In Papers in Laboratory Phonology II: Gesture, Segment, Prosody (G.J. Docherty & D. Robert Ladd, Eds.), 68-86. Cambridge University Press: Cambridge.

Browman, C.P. & Goldstein, L. (1990) *"Targetless" schwa: An articulatory analysis,* Haskins Laboratories Status Report on Speech Research, SR-101/102, 194-219.

De Jong, K. (1991) *An articulatory study of consonant-induced vowel duration changes in English,* Phonetica, 48, 1-17.

Edwards, J., Beckman, M.E. & Fletcher, J. (1991) *The articulatory kinematics of final lengthening,* Journal of the Acoustical Society of America, 89, 369-382.

Gay, T. (1981) *Mechanisms in the control of speech rate,* Phonetica, 38, 148-158.

Harrington, J., Fletcher, J. & Roberts, C. (submitted) *Coarticulation and the accented/ unaccented distinction: evidence from jaw movement data,* manuscript submitted for publication to the Journal of Phonetics.

Kuehn, D. & Moll, K. (1976) *A cineradiographic study of VC and CV articulatory velocities,* Journal of Phonetics, 4, 303-320.

Lindblom, B. (1963) *Spectrographic study of vowel reduction,* Journal of the Acoustical Society of America, 35, 1773-1781.

Salzman, E. (1991) *The task dynamic model in speech production.* In Speech Motor Control and Stuttering, pp. 37-52 (Peters, Hulstijn and Starkweather, Eds. Elsevier Science Publications)

Wieneke, G., Janssen, P. & Belderbos, H. (1987) *The influence of speaking rate in the duration of jaw movements,* Journal of Phonetics, 15, 111-126.