Analyzing speech movement variability from audio recordings

Frits van Brenk[1], Anja Lowit[1]
[1] Division of Speech and Language Therapy, University of Strathclyde, Glasgow, UK.

Introduction

Producing fluent and intelligible speech involves quick, precise and coordinated movements of articulators. Developmental and acquired speech motor disorders often result in an instable gestural organization of speech movements, and thus more variable speech performance. In addition, research suggest that a number of environmental factors can influence articulatory movements, including speech rate modifications, linguistic and cognitive complexity of the speech task and concurrent motor tasks. The possible interaction of internal and external factors that influence speech motor control might give more insight in the nature of different speech disorders. The spatiotemporal index (STI) (Smith et al., 1995) and Functional Data Analysis (FDA) (Ramsay et al., 1996) are techniques that have been shown to be succesfull in investigating variability of kinematic movements obtained by lip displacement tracking, but may also be applied to other dimensions of speech, including amplitude envelopes and pitch and formant tracks. This pilot study presents some first results on STI and FDA variability measures applied to audio recordings of young, healthy adults. The study aims to investigate whether variability measures obtained from acoustic data reflect those from kinematic data.

Methodology

Particpants included seventeen healthy native Scottish speakers, 13 females and 4 males, age range 18 to 45 years (mean = 27.2 years, SD = 8,6 years). They repeated the test phrase "Tony knew you were lying in bed" around 20 times as exactly as possible in six conditions as follows: (1) habitual speech rate (baseline condition), (2 & 3) slow and fast rate (modified rate condition), (4 & 5) habitual rate at increased sentence length with and without increased syntactic complexity (sentence length condition), and (6) habitual rate whilst performing a simultaneous drawing task (dual task condition).

The digital audio recordings were processed in Speech Filing System (SFS v4.7) software. The amplitude envelope (E), fundamental frequency (F0), first formant (F1) and second formant (F2) tracks were extracted. The beginning and end of the voiced parts of each repetition were marked in the oscillogram, and served as the analysis window for the variability calculation. The envelopes were resampled at 8000 Hz and filtered using a 6th order Butterworth filter, with a cut-off frequency of 10 Hz, and were then normalized by calculating the z-scores. To calculate the STI, the envelopes were linearly stretched and squeezed in order to fit them all onto the same time frame. Their mean and SD were calculated to obtain the STI. For the FDA analysis, the tracks were non-linearly stretched, allowing a separate calculation of spatial as well as temporal variability.

Results

The different speaking conditions were compared by one-way ANOVA. Only significant results ($p < .05$) will be reported.

Articulation rate Statistical results indicate that participants changed articulation rate significantly between the three different rate conditions (slow, habitual and fast). Sentence complexity influenced articulation rate, and post-hoc analysis revealed that articulation rate was significantly slower in the longer and more complexer sentences, compared to the baseline condition. No difference in articulation rate was found between the longer sentences with and without increased syntactic complexity. The dual task condition did not have influence on articulation rate.

Modified rate The STI of the F1 track was higher in the slow condition, compared to habitual and fast condition. Spatial variability of the amplitude envelope was higher in the slow condition, compared to habitual and fast condition. Spatial variability of F1 track was higher in the slow condition, compared to the fast condition. Temporal variability of the F2 track was higher in the fast condition, compared to the habitual and slow speech rate conditions.

Sentence length condition Both spatial variability and temporal variability of the F0 track were higher in both increased complexity conditions, compared to the baseline condition. The F0-STI was significantly higher in both increased complexity sentences, compared to the base sentence.

Dual task Temporal variability of the F0 track was higher during the dual task, compared to the base condition.

Discussion

The speakers were successful in changing articulation rate across different rate conditions. In addition, articulation rate was slower in the more complex sentences.

In relation to variability measures the study showed that modifying speaking rate from habitual to fast or slow resulted in an increase in STI as well as spatial and temporal variability for some of the speech dimensions, reflecting the results on lower lip movement variability found by Smith et al. (1995). However, increasing sentence length and complexity also resulted in an increase in variability, contradicting earlier results on lower lip variability (Kleinow and Smith, 2000). The increase in variability found during the dual task condition reflects the findings of Dromey and Benson (2003).

Some methodological issues have been encountered, including interruptions of continuous voicing throughout the speech phrase and inconsistent formant estimations. Possible solutions include interpolation algorithms or manual redrawing of formant tracks.

Despite the difference in some of the results to previously published literature, this study has demonstrated that variability measures obtained from audio recordings can reflect changes in motor control in a similar way to the kinematic measures reported so far. This method can be a promising alternative to the invasive and complex nature of direct measurements of kinematic movements, allowing for a greater freedom in experimental set-up. The current results serve as a good basis to compare and extend to populations with speakers with motor speech disorders.

References

Dromey, C. and Benson, A. (2003). Effects of concurrent motor, linguistic, or cognitive tasks on speech motor performance. *Journal of Speech, Language and Hearing Research*, 46(5):1234.

Kleinow, J. and Smith, A. (2000). Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter. *Journal of Speech, Language, and Hearing Research*, 43:548–559.

Ramsay, J. O., Munhall, K. G., Gracco, V. L., and Ostry, D. J. (1996). Functional data analyses of lip motion. *Journal of the Acoustical Society of America*, 99(6):3718–3727.

Smith, A., Goffman, L., Zelaznik, H. N., Ying, G., and McGillem, C. (1995). Spatiotemporal stability and patterning of speech movement sequences. *Experimental Brian Research*, 104:493–501.