KINEMATIC AND ACOUSTIC ANALYSIS OF ARTICULATORY GESTURE PHASING

Gary Weismer and Ray D. Kent
Waisman Center & Dept. Communicative Disorders
University of Wisconsin-Madison

ABSTRACT

Kinematic and acoustic measures were made to explore the relationship between the two. Results indicate that measures of F2 onset and offset, previously hypothesized to be likely indices of extent of gesture overlap, can in some cases be predicted with good reliability from kinematic measures.

INTRODUCTION

In two recent papers (Weismer, Tjaden, & Kent, in press a,b) we have argued that acoustic measures can serve as indices of articulatory gesture overlap. Measures of onset and offset frequencies of the second formant (F2) in CV and VC sequences have, in certain situations, been shown to vary in a manner consistent with a logical analysis of gesture overlap. However, in the absence of data relating aspects of actual articulatory gestures to the acoustic output of the vocal tract, the relative goodness of these inferences will remain unknown. The purpose of this paper is therefore to report some exploratory coanalysis of x-ray microbeam data and speech acoustic measures from selected speakers and utterances. In particular, we were interested in the extent to which certain kinematic measures could be used to predict F2 onset and offset measures. The relative goodness of these predictions should point to some guidelines for the inference of articulatory behavior from acoustic measures.

METHODS

Kinematic and acoustic data were collected as part of the x-ray microbeam data base project (Westbury, 1994). The data base consists of over fifty speakers who produced a common speech sample of material ranging from simple syllables to a relatively lengthy reading passage.

Subjects

The completed analysis will be based on ten speakers. In the present report, data are presented for two speakers, including a male aged 28 years (JW7) and a female aged 20 years (JW31). Both subjects had normal orofacial structures, spoke a dialect generally described as Greater American Midwest (one speaker grew up in Iowa, the other Wisconsin), and reported no history of speech or language problems.

Speech Sample

Kinematic and acoustic measures were obtained from the sequence /ubIg/ in the utterance, *The other one is too big*. Between 15 and 20 repetitions of this utterance were produced by each subject, these repetitions including several at slower-than-normal and faster-than-normal speaking rates. The rate variation was desirable for the present analyses, because of the presumed effect of rate on the extent of gesture overlap (see, for example, Munhall & Lofqvist, 1992; Tjaden, Weismer, & Kent, 1994) and thus on variation in F2 onset and offset.

Data Collection

X-ray microbeam data for the two subjects reported herein were collected using the standard array of 11 pellet locations used throughout the data base project. For the purposes of the present project, attention was focussed on the three pellets attached at the mid-ventral, mid-dorsal, and dorsal tongue locations (see Westbury, 1994, p. 39). The approximate distances of these pellets from the tongue apex (measured along the surface of the extended tongue) were 25, 44, and 60 mm, respectively (the dorsal pellet was not tracked in subject JW7 due to technical problems at the time of data collection). The speech acoustic signal was digitized and stored synchronously with the pellet position histories. Complete technical details concerning data collection for the x-ray microbeam data base project are provided in Westbury (1994).

Measures

The measures taken in the current investigation are summarized in Fig. 1, which shows pellet time histories in the *x*

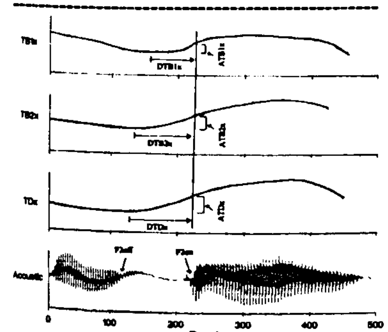

Figure 1. Graphic summary of measures.

dimension (antero-posterior) for the mid-ventral (TB1x), mid-dorsal (TB2x), and dorsal (TDx) markers, as well as the synchronized acoustic waveform of the /ubIg/ sequence. Downward pellet time histories reflect posterior tongue movement, and upward histories forward tongue movements. The pattern shown in Fig. 1, where all pellets describe a U-shaped path throughout the sequence of interest, characterized each utterance analyzed for the present report. This pattern is consistent with the phonetic expectation of tongue backing for the /u/ and then fronting for the /I/. The point in time at which the movement began to be directed forward usually occurred within the /b/ closure interval, but sometimes occurred prior to the acoustic evidence of the onset of the closure interval. Using this point in time, we defined DTB1x, DTB2x, and DTDx as the temporal interval between the onset of the forward-directed motion and the first glottal pulse of the vowel /I/ in *big*. The first glottal pulse for /I/ was chosen as the termination point for this temporal measure because it serves as the location of the F2on measure we have evaluated as an index of gesture overlap. We reasoned that variations in the duration of this interval of tongue movement might be correlated with F2on, if earlier onsets of the forward-directed tongue movement could be interpreted as greater overlap between the lingual and labial gestures for this phonetic sequence.

We also measured the amount of forward movement throughout these temporal intervals, indicated in Fig. 1 as ATB1x, ATB2x, and ATDx. We reasoned that variations of extent of forward movement may also be correlated with variations in F2on.

The acoustic measures taken included the F2on value measured above, the F2off value of /u/, measured at the last glottal pulse preceding the closure interval, the /b/ closure duration and the /b/ VOT. Formant measures were made using a combination of LPC spectra and cursor placement on digital spectrograms; temporal measures were obtained from the combined digital spectrogram/waveform display.

RESULTS AND DISCUSSION

Averaged data for selected measures described above are presented in Table 1.

Table 1. Averaged data for measures.

	DTB1x	DTB2x	ATB2x	F2on	F2off
JW31	56	104	1.39	1886	1311
JW7	103	103	7.20	1468	903

These data indicate that JW31's tongue displacements (in mm) in the *x* dimension were substantially smaller than JW7's tongue displacements. The two subjects showed very different temporal intervals (in msec) for TB1x, but essentially identical intervals for TB2x (no comparison could be made for TDx because of the technical problem noted above). Closure duration and VOT were quite similar for the two subjects, and the differences in F2 measures can probably be attributed largely to gender-based differences in vocal tract size.

Table 2 reports the significant results from exploratory analyses in which hypothesized acoustic indices of gesture overlap, F2off for the /u/ in *too* and F2on for the /I/ in *big*, were regressed on the temporal and displacement measures obtained from the pellet time histories. Because of high intercorrelations between either temporal and displacement measures, or displacement measures for two different pellets, some of these effects are mutually redundant. For example, effects *1* & *2* for

Table 2. Significant regression effects.

JW31
1. F2on=1994-1.95*DTB1x, p=.018, R^2= 29%
2. F2on=1961-117*ATB1x, p=.019, R^2= 29%
3. F2off=1445-208*ATB1x, p=.003, R^2= 45%
4. F2on=1813-2.25*OTB1x, p=.010, R^2= 34%

JW7
1. F2on=1390+11.1*ATB1x, p=.014, R^2= 33%
2. F2on=1382+12.0*ATB2x, p=.020, R^2=30%
3. F2off=1315-4.02*DTB1x, p<.001, R^2=67%
4. F2off=1124-31.5*ATB1x, p<.001, R^2=78%
5. F2off=1161-35.9*ATB2x, p<.001, R^2=80%

both JW31 and JW7, and effects *3*, *4*, and *5* for JW7 are alternate expressions of the same phenomena because of the kinds of interdependencies cited above. With this in mind, a general summary of the regression analyses is as follows. First, F2on appears to be related significantly to displacement between the onset of forward movement and the first glottal pulse of /I/ (ATBx), albeit in different directions for the two subjects. For JW31 larger ATBx was related to lower F2on, whereas the opposite was the case for JW7. Intuitively, it seems as if the latter pattern would be more consistent with the idea that earlier onsets of forward movement (and the associated larger displacements over the DTBx interval) would reflect greater gesture overlap between the lingual and labial gestures for this sequence. Second, F2off of /u/ is strongly related to the ATBx magnitude for both subjects, such that larger ATBx was associated with lower F2off. Figure 2 shows these effects as scatterplots for both subjects.

None of the analyses reported above was based on a direct measure of gesture overlap. To explore the predictive utility of a measure that might more directly express the interarticulatory phasing of lingual and labial gestures, we subtracted, token by token, the closure interval + VOT from

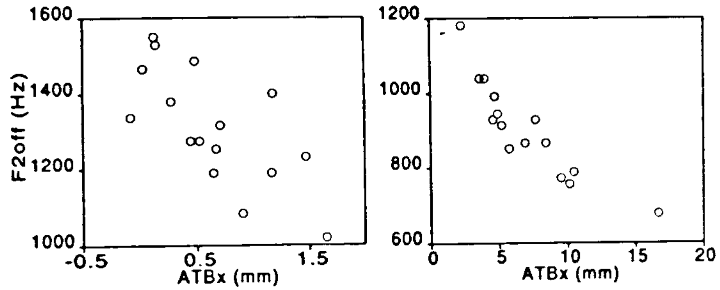

Fig. 2. Scatterplots of F2off-ATBx effects for subjects JW31 (left) and JW7 (right).

DTBx which produced a measure of the initiation of forward-directed tongue movement toward /I/ relative to the acoustically-inferred onset of labial closure for /b/. This interval was, with a single exception among 16 repetitions, always negative for JW31, indicating that the forward movement began after the labial closure; as seen in Table 2 the F2on was predicted significantly from this measure of overlap (OTBx), the nature of the effect being that the less negative the measure of overlap the lower the F2on. This is exactly what we would expect if F2on was to be used as a reliable indicator of gestural overlap, because less negative values of OTBx would imply that the forward directed tongue movement was beginning earlier in time relative to the onset of labial closure for /b/ (i.e., that there was greater gestural overlap between the labial and lingual gestures). Unfortunately, a corresponding analysis of JW7's data failed to reveal a significant function.

The data reported here show some promise in establishing reasonable links between kinematic and acoustic measures. Obviously the extent of speaker variation must be known, and exploration of better measures of interarticulator phasing undertaken, before straightforward articulatory inferences can be made from acoustic measures.

REFERENCES

MUNHALL, K., & LOFQVIST, A. (1992). Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics, 20*, 111-126.

TJADEN, K., WEISMER, G., & KENT, R.D. (1994). *Evidence of gestural overlap across speaking rate: Acoustic data.*. Paper presented at the 128th meeting of the Acoustical Society of America, Austin, TX, USA.

WEISMER, G., TJADEN, K., & KENT, R.D.(in press). Can articulatory behavior in motor speech disorders be accounted for by theories of normal speech production? *Journal of Phonetics, 23*.

WEISMER, G., TJADEN, K., & KENT, R.D.(in press). Speech production theory and articulatory behavior in motor speech disorders. In F. Bell-Berti and L.J. Raphael (Eds.), *Studies in speech production: a festschrift for Katherine Safford Harris*. New York: American Institute of Physics.

WESTBURY, J.R. (1994). X-ray Microbeam Speech Production Database User's Handbook, Version 1.0 (131 pages). University of Wisconsin-Madison.