MAGNETOMETER AND X-RAY MICROBEAM COMPARISON

Dani Byrd*†, Catherine P. Browman*, Louis Goldstein*‡, and Douglas N. Honorof*‡
*Haskins Laboratories, USA, †University of Southern California, USA, ‡Yale University, USA

ABSTRACT

This article concerns replicability of past results and consistency in data collection using X-ray microbeam and magnetometry, two prominent instruments for speech movement tracking. The same speaker was recorded on both instruments reading identical stimuli and using parallel experimental procedures and the same physician for transducer placement. Data from two X-ray microbeam (XRMB) runs were collected on the same day (transducers re-placed for the second run); 27 months later the analogous data were collected with the Electromagnetic Midsagittal Articulometer (EMMA) system. Vertical movement of the lower lip, tongue tip, and tongue body was analyzed. Extremum positions and peak velocities were identified, and movement displacements and durations calculated. The correlation of these measures between the EMMA and XRMB sessions is very high, almost as high as that between the two XRMB sessions. Subsequent ANOVA confirm that the results obtained using EMMA do not differ substantially from those using X-ray microbeam.

1. INTRODUCTION

The establishment of consensus in science requires the replicability of empirical results. Therefore, it is crucial that new data collection systems be evaluated in functionally identical experimental settings against older systems before (inevitable) comparisons are drawn between studies using the differing systems. Much articulatory movement data have been archived from the X-ray microbeam facility (e.g., [1, 2, 3, 4]). Currently, however, magnetometer systems such as the Electromagnetic Midsagital Articulometer are the predominant tool for tracking articulator movement. In the interests of replicability of past results and consistency in data collection for ongoing research programs drawing data from both sources. There are, for example, several factors that may interfere with quantitative consistency between experimental data collected with EMMA and with XRMB. For instance, off-midline movements of the transducers including tilt are known to occur during the articulator movement of speakers and are an accuracy concern for magnetometer systems [5].

Both the electromagnetic midsagittal articulometer and the X-ray microbeam are tracking devices that record movement of midsagittal points on the articulators including the tongue, lips, and points on the jaw. Our work considers data from the Electromagnetic Midsagittal Articulometer, or EMMA, magnetometer system and the NIH X-ray microbeam system (henceforth XRMB) at the University of Wisconsin. The technical specifications of the EMMA system are outlined in [5] (see also [6]) and those of the XRMB system can be found in [7]. (There are other alternating magnetic field systems for speech research; see for example [8]. We consider only the EMMA system here.) As part of an ongoing research project at Haskins Laboratories, identical speech stimuli from the same speaker were collected with both the EMMA and XRMB systems in order to evaluate consistency between the instruments. Further work is underway to consider complementary MRI data as well.

2. METHOD

A female speaker who grew up in Chicago, Illinois, was recorded speaking the same utterances using both EMMA and XRMB. Data from two XRMB sessions were collected on the same day: the first session in the morning and the second in the afternoon. The pellets attached to the articulators were re-placed for the second run. Twenty-seven months later, the analogous EMMA data were collected.

Three steps were taken to ensure maximum agreement in transducer placement across the three sessions. (1) The same physician placed the movement transducers in each session, following the same protocol. (2) To aid with transducer placement reliability, a "touch bar" was designed to provide the subject with a fixed target for tongue protrusion. The bar was attached by a splint to one of her lower molars, and the subject could then protrude her tongue until the tip just touched the bar. The subject held her tongue in this position during placement, allowing measurements of transducer position to be made with the tongue consistently stretched to a fixed length. (3) During the first XRMB session, photographs of the subject's tongue were taken after pellet placement and with the tongue protruded to the touch bar, and these were used as a guide by the physician in subsequent sessions.

Measured distances of the tongue transducers from the tip of the tongue (when touching the bar) for the three sessions are shown in Table I. As can be seen, the placements agree within one mm across all three sessions. For both instruments, lip transducers were placed immediately above the vermilion border for the upper lip and below it for the lower lip. Additional transducers not examined here were placed on the nose, (upper and lower) teeth, and in the XRMB sessions, one was attached by string from the rearmost fixed pellet and rested on the back of the tongue. The XRMB pellets attached to the articulators were 2.5 mm in diameter, while the EMMA movement transducer coils, including the glue used to adhere them, were approximately 2.0 x 4.1 x 1.8 mm. (Currently, smaller coils are available.) For this study, signals for three of the transducers are examined: the lower lip, the tongue tip, and the second of the three transducers on the tongue body (TB2).

transducer / session	TT (tongue tip)	TB1 (tongue body 1)	TB2 (tongue body 2)	TD (tongue dorsum)
EMMA	7	23	41	66
XRMB 1	7	23	41	65
XRMB 2	6	22	40	65

Table 1. Distance (in mm) from tip of tongue to four transducers on the surface of the tongue.

For both instrument types, the orientation of the subject's occlusal plane within the measurement field was determined by having the subject bite down on a plate, the surface of which had transducers attached to it. The measured positions of the transducers on the bite plate were used to define the occlusal plane, and the acquired data were rotated to a new coordinate system, in which the horizontal axis is aligned with the occlusal plane. For both instruments, the y-origin of the transformed coordinate system is on the bite plate line. However, because of a minor difference in the bite plate procedures (see section 3.2), there may be a small vertical shift between data from the two instruments. In particular, before the first XRMB session, the subject produced a dental impression, which was then attached to the bite plate. For recording the occlusal plane angle, the subject held the bite plate in her mouth by placing her teeth in the dental impression. In the EMMA session, no dental impression was used, and the subject held the bite plate directly between her upper and lower teeth.

The XRMB samples were acquired at an effective rate of 80 Hz for the lip and tongue body pellets and 160 Hz for the tongue tip pellet. Following correction for head movement and rotation to the occlusal plane (see [1]), the lip and tongue body data were subject to 3-point smoothing (resulting in an effective 16.35 Hz low pass filter cutoff) and the tongue tip data to 5-point smoothing (20.04 Hz cutoff) with triangular filters. The data were then interpolated (linearly) to a nominal 200 Hz sampling rate. The EMMA data were sampled at 625 Hz after low-pass filtering at 300 Hz before voltage-to-distance conversion. Following correction for head movement (using the nose and maxillary reference transducers) and rotation to the occlusal plane, the data were subject to 21-point smoothing (16.5 Hz cutoff) by a triangular filter.

The utterances analyzed for this comparison study were "It's a [pV'CVp] again" where the vowels (V-V) were [a-a], [a-i], [i-a], [i-i], and [u-u] and the consonants (C) were [p], [t], and [k]. Vowel contexts (V-V) were blocked and recorded with the consonants in a rotating order of [p], [t], and [k]. The utterances were recorded as part of a larger study that included other material not analyzed here. The night before the EMMA session, the subject listened to the tokens she had produced in the first XRMB session, to remind her of the overall rate and style that she had employed.

In the XRMB sessions three repetitions of each utterance were recorded. Any repetitions with mistrackings or perceived speech errors were excluded from analysis. In the EMMA session, four to five repetitions were recorded; any with speech errors were excluded. In the analyses below, only parallel repetitions were compared. That is, repetition one of an utterance in the first XRMB session was compared only to repetition one of that utterance in the second XRMB session and in the EMMA session. Utterances for which parallel repetitions in all three conditions did not exist (i.e., repetitions four and five in the EMMA session) were not included in the analysis. (The one exception to this was the substitution of EMMA repetition four of [pukup] for repetition three, which had a speech error.)

Vertical movement data for the lower lip, tongue tip, and one tongue body transducer were considered for the consonant and the (same) tongue body transducer for the adjacent vowels. An automatic peak-picking algorithm (with a spatial noise level of .2 mm) was used to record the absolute positions (i.e., extremum positions in mm from the occlusal plane) and the times at which these extrema occurred. (We will refer to these extrema as peaks in the case of local maxima and valleys in the case of local minima.) The extrema of interest for the vowels and consonants for the different utterance types and articulators were peaks and valleys that indicated the beginnings and ends of the relevant opening and closing movements. The consonant-related events included five lower lip points for $[p_1V_x'p_2V_yp_3]$ utterances—the first peak for $[p_1]$, the valley during Vx, the peak for $[p_2]$, the valley during V_y, and the peak for $[p_3]$. For [pV'tVp] utterances, three tongue tip points were considered—the initial valley, followed by the peak during the closure, and the final valley. For [pV'kVp] three tongue body extremum positions —valley, peak, valley—were considered for symmetrical vowel environments. In non-symmetrical vowel environments, two tongue body positions were considered—peak and valley in [ika] utterances and valley and peak in [aki] utterances. Vowel-related events were measured in [p] and [t] tokens in non-symmetrical vowel environments. Tongue body peaks during [i] and valleys during [a] were considered. Examples of the utterance "It's a [pi'pap] again" with the extremum positions marked in a token from the XRMB1 session and the EMMA session are shown in Figure 1.

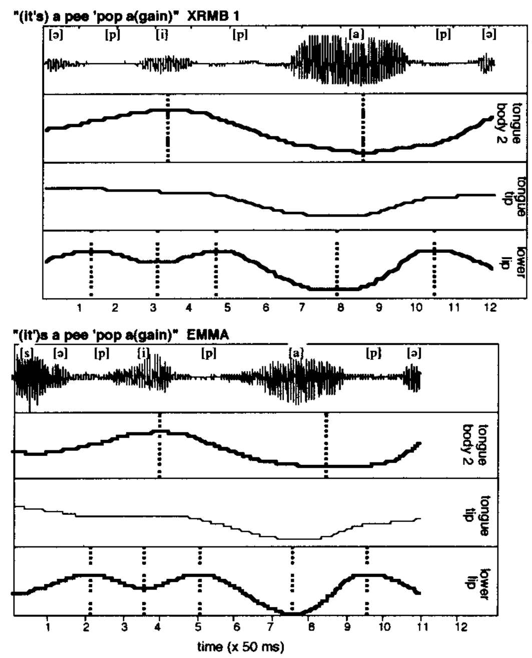

Figure 1. Examples of the utterance "It's a [pi'pap] again" with the extrema positions marked in a token from the XRMB session 1 (XRMB1) and in one from the EMMA session. The two are aligned at the first lower lip peak. Relevant time functions for this utterance are shown in bold.

After the extremum positions and times were recorded, movement displacement and duration were calculated by taking the absolute values of the difference in time and spatial position between *extremum position*$_n$ and *extremum position*$_{(n+1)}$. For example, if extrema for a valley-peak-valley series were measured for a [t], three position values were obtained and two duration and two displacement measures ($valley_1$ to peak, and peak to $valley_2$). Finally, peak velocities were obtained for each opening and closing movement by the following procedure. Velocity signals were smoothed using the same windows described above for the movement signals. A noise level of 5% of the overall velocity range for each transducer dimension was used in picking the velocity peaks. If multiple velocity peaks were found for a single closing or opening movement, the largest velocity value in the correct direction was used. A total of 432 absolute position measures, 302 duration and displacement measures, and 300 peak velocity measures were obtained for further statistical analyses.

Two types of analysis were conducted. First, the Pearson product-moment correlation was calculated using a three-way correlation matrix and listwise deletion (i.e. if a value was missing for any cell in the correlation matrix, that token was excluded from all cells in the matrix) to determine how well the correlation of measurements made with different instruments (EMMA vs. XRMB1/2) compares to that of measurements from two sessions recorded with the same instrument (XRMB1 vs. XRMB2). These matrices were calculated separately for each articulator. Second, analyses of variance were used to test for main effects of session (EMMA vs. XRMB1 vs. XRMB2) and any interactions of session with particular articulators or vocalic contexts. All analyses were conducted for the dependent variables of position, displacement (unsigned), duration, and peak velocity (for which absolute values were used).

3. RESULTS

3.1. Correlation

Scatter plots of position and duration measures for EMMA versus XRMB1 can be found in Figure 2A and 2B. Those for XRMB1 and XRMB2 are in Figure 2C and 2D.

For all four variables—position, displacement, duration, and peak velocity—and for all three articulators the correlations between each XRMB and the EMMA session were very high and almost as high as the correlations between the two XRMB sessions. Recall that the XRMB sessions were recorded hours apart, while the XRMB and EMMA sessions were recorded 27 months apart. The r values for the correlations of position, displacement, duration and peak velocity between sessions for each articulator can be found in Table II. Nearly all correlations had a $p < .0001$. Two cells had correlations with $p<.01$. The only exception to this was the lack of correlation between any of the sessions for the duration of the tongue tip movement.

The significant EMMA/XRMB correlations range from .9818 in the position values to .4613 in the duration values. The XRMB1/XRMB2 correlations range from .9852 for position to .5638 for peak velocity. One possible explanation for the lower duration correlations is that there are (non-uniform) differences in speaking rate between the sessions. Accordingly, duration would be expected to vary most of the four measures. Finally, of the 22 significant EMMA/XRMB correlations, 11 were even higher than the parallel correlations between the two XRMB sessions.

3.2. Analyses of Variance

Three-factor full-interaction ANOVAs testing for effects of session (3-level: EMMA, XRMB1 and XRMB2), vocalic context (5 level: aCa, aCi, iCa, uCu), and articulator (3-level: lower lip, tongue tip, and tongue body) on position, displacment, duration

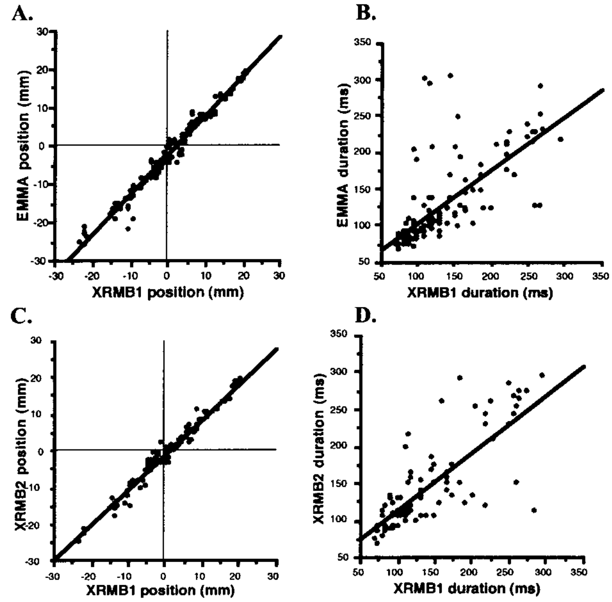

Figure 2. Panel A: position values for the XRMB1 vs. the EMMA session. Panel B: duration values for the XRMB1 vs. the EMMA session. Panel C: position values for XRMB1 vs. the XRMB2 session. Panel D: duration values for the XRMB1 vs. the XRMB2 session. Regression lines for a linear fit are shown.

	position	displacement	duration	peak velocity
XRMB1 & EMMA				
lower lip	.9525**	.8127**	.6171**	.7776**
tongue tip	.9817**	.9584**	.3140	.8731**
tongue body	.9897**	.8499**	.6388**	.6984**
XRMB2 & EMMA				
lower lip	.9656**	.8665**	.4613*	.8401**
tongue tip	.9818**	.9351**	.2361	.9510**
tongue body	.9848**	.9241**	.7503**	.8777**
XRMB1 & XRMB2				
lower lip	.9785**	.9339**	.6644**	.9139**
tongue tip	.9782**	.9140**	.2502	.8129**
tongue body	.9852**	.8288**	.5955*	.5638*
	N=60, 46, 32	N=48, 25, 21	N=48, 25, 19	N=48, 27, 22

Table 2. r-values for the correlation of the EMMA and two XRMB sessions; where N is the number of observations shared by all three sessions for each articulator—lip, tongue tip, and tongue body. ($^{**}p < .0001$, $^{*}p < .01$).

and peak velocity were conducted to determine if the EMMA data differed from the XRMB data. Means and standard deviations for these measures are given in Table III. (Note that Table III does not include a breakdown by vowel context. Therefore, the standard deviations shown are much higher than those for a further breakdown by vowel context, and it is the latter that are used in the computation of the error term for the ANOVA.)

	position *(mm)*	displacement *(mm)*	duration *(ms)*	peak velocity *(mm/s)*
EMMA				
lower lip	-10.504 (5.561)	9.6154 (4.5553)	91.8077 (14.9326)	168.8776 (71.6115)
tongue tip	0.465 (6.149)	10.9583 (4.2475)	164.6250 (74.8865)	164.2374 (76.1182)
tongue body	10.133 (6.732)	11.0357 (4.3161)	185.8929 (49.5516)	126.7271 (44.4975)
XRMB 1				
lower lip	-7.2278 (5.6577)	9.5885 (4.8529)	102.3077 (23.5660)	144.9963 (57.9854)
tongue tip	3.4825 (6.1340)	10.8583 (4.4137)	143.1250 (47.2232)	165.6624 (66.8590)
tongue body	11.6018 (7.0424)	12.8192 (3.6322)	213.2000 (49.0298)	116.2879 (41.2403)
XRMB 2				
lower lip	-7.3960 (6.0719)	10.6082 (5.1266)	111.2500 (25.0850)	152.6237 (68.5298)
tongue tip	1.7002 (6.7386)	11.5095 (5.2295)	140.7143 (34.3979)	172.2895 (74.0941)
tongue body	10.0722 (7.0365)	12.4269 (3.7273)	216.2963 (58.1707)	115.0355 (35.9258)

Table 3. Means and standard deviations for position, displacement, duration, and peak velocity for the EMMA and two XRMB sessions.

ANOVA found no main effect of session on displacement ($F(2,257) = 1.1$), duration ($F(2,256) = 1.07$) or peak velocity ($F(2,225) = .801$). There was a significant effect of session on position $F(2,387)=5.085$, $p<.0066$) distinguishing the EMMA session from the other two sessions (determined by a Fisher's PLSD post-hoc test ($p<.05$)). The 2.27 mm (averaged) difference is likely due to the difference in procedure for determining the occlusal plane described above, and therefore is uninteresting. The presence of the dental impression in the case of the XRMB procedure would effectively place the subject's maxilla slightly higher above the bite plate than in the EMMA procedure, and indeed it is the XRMB data that show higher positions with respect to the occlusal plane. Finally, there was a significant interaction of session and articulator on duration ($F(4,256) = 3.483$, $p = .0093$). This was due to shorter durations in the EMMA session for the tongue body articulator. No other main effects or interactions were significant—in particular, session had no systematic effect on displacement, duration, or peak velocity.

Another way of assessing the comparability of the different techniques is to examine the data for an expected phonetic effect, so as to confirm that both techniques show the effect and that there is no difference in the robustness of the effect. One candidate is the effect of vowel context on displacement of the consonant articulators into and out of a consonant constriction. The displacement of the receivers most relevant to the consonant constriction can be expected to be greater when the consonant is produced in the context of a low vowel than in the context of a high vowel. To test this, tokens of [pV'pVp] and [pV'tVp] were analyzed ([pV'kVp] tokens were not analyzed because consonant and vowel constrictions are conflated in the same receiver in this case). A three-factor full-interaction ANOVA testing for effects of session, vocalic context, and articulator (2 level: lower lip & tongue tip) on displacement found the expected main effect of context such that the low vowel contexts cause greater displacement for the consonant articulator ($F(4,192) = 24.628$, $p < .0001$). Importantly, however, there was no interaction of session and environment ($F = .359$) or articulator ($F = .303$).

4. CONCLUSION

This report has addressed concerns of replicability and consistency that have arisen from the use of two different instruments in collecting articulatory movement data. We have examined the degree to which data acquired with the EMMA and the NIH X-ray microbeam (XRMB) movement transduction systems are consistent. In conclusion, high correlations between position, displacement, duration, and peak velocity data were obtained between sessions recorded 27 months apart with the EMMA and the X-ray microbeam instruments. The correlation of these measures between the EMMA and XRMB runs is very high and is almost as high as that between the two XRMB runs. Furthermore, analyses of variance find no substantive differences between the instruments. These data demonstrate that extremely similar kinematic results can be obtained from recordings of a speaker made at different times and with different instruments—given consistency in subject behavior, in data elicitation procedures, and in movement transducer placement. Specifically, this unique data set allows us to conclude confidently that differences occurring in data collected with the EMMA system and the X-ray microbeam system will be due to factors other than the type of movement transducer system used.

ACKNOWLEDGMENTS

The authors thank Catherine Best, Vince Gracco, Katherine Harris, Kiyoshi Oshima, John Westbury, Ed Wiley, and the staff of the NIH microbeam system at the University of Wisconsin, especially Carl Johnson, for help of various sorts. The authors are very grateful to two anonymous reviewers, Freddie Bell-Berti, Anders Löfqvist, Joe Perkell, and Pascal H. H. M. van Lieshout for their comments on the manuscript. This research was supported by NIH grants DC-00121 and DC-00016 to Haskins Laboratories.

REFERENCES

[1] Westbury, J.R. 1994. *X-ray microbeam speech production database user's handbook, version 1.0*, Madison, WI.

[2] Sproat, R. and Fujimura, O. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21. 291-312.

[3] Johnson, K., Ladefoged, P. & Lindau, M. 1993. Individual differences in vowel production. *J. Acoust. Soc. of Am.*, 94,2, pt. 1. 701-714.

[4] de Jong, K. 1991. An articulatory study of consonant-induced vowel duration changes in English, *Phonetica* 48. 1-17.

[5] Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. 1992. Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *J. Acoust. Soc. of Am.*, 92, 3078-3096.

[6] Gracco, V.L. and Nye, P.W. 1993. Magnetometry in speech articulation research: Some misadventures on the road to enlightenment. *Forschungsberichte des Institut für Phonetic und Sprachliche Kommunikation der Universität München (FIPKM)* 31, 91-104.

[7] Nadler, R.D., Abbs, J.H., and Fujimura, O. 1987. Speech movement research using the new X-ray microbeam system. *Proceedings of the XIth International Congress of Phonetic Sciences*, 1, 221-224.

[8] Schönle, P.W. 1988. *Elektromagnetische Artikulographie*. Berlin, Springer.