1-1-2013

# Vowel Production in Mandarin Accented English and American English: Kinematic and Acoustic Data from the Marquette University Mandarin Accented English Corpus

An Ji
*Marquette University*

Jeffrey J. Berry
*Marquette University*, jeffrey.berry@marquette.edu

Michael T. Johnson
*Marquette University*, michael.johnson@marquette.edu

---

Published version. Published part of the proceedings of the conference, *ICA 2013 Montreal*, 2013. DOI. © 2013 Acoustical Society of America. Used with permission.

Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

## ICA 2013 Montreal
## Montreal, Canada
## 2 - 7 June 2013

## Speech Communication
## Session 4pSCb: Production and Perception I: Beyond the Speech Segment (Poster Session)

# 4pSCb40. Vowel production in Mandarin accented English and American English: Kinematic and acoustic data from the Marquette University Mandarin accented English corpus

**An Ji, Jeffrey J. Berry* and Michael T. Johnson**

***Corresponding author's address: Speech Pathology & Audiology, Marquette University, Milwaukee, WI 53201-1881, jeffrey.berry@marquette.edu**

Few electromagnetic articulography (EMA) datasets are publicly available, and none have focused systematically on non-native accented speech. We introduce a kinematic-acoustic database of speech from 40 (gender and dialect balanced) participants producing upper-Midwestern American English (AE) L1 or Mandarin Accented English (MAE) L2 (Beijing or Shanghai dialect base). The Marquette University EMA-MAE corpus will be released publicly to help advance research in areas such as pronunciation modeling, acoustic-articulatory inversion, L1-L2 comparisons, pronunciation error detection, and accent modification training. EMA data were collected at a 400 Hz sampling rate with synchronous audio using the NDI Wave System. Articulatory sensors were placed on the midsagittal lips, lower incisors, and tongue blade and dorsum, as well as on the lip corner and lateral tongue body. Sensors provide five degree-of-freedom measurements including three-dimensional sensor position and two-dimensional orientation (pitch and roll). In the current work we analyze kinematic and acoustic variability between L1 and L2 vowels. We address the hypothesis that MAE is characterized by larger differences in the articulation of back vowels than front vowels and smaller vowel spaces compared to AE. The current results provide a seminal comparison of the kinematics and acoustics of vowel production between MAE and AE speakers.

Published by the Acoustical Society of America through the American Institute of Physics

**[DOI: 10.1121/1.4800290]**
**Received 22 Jan 2013; published 2 Jun 2013**

# INTRODUCTION

Methods of pronunciation assessment and strategies for the modification of the pronunciation differences of second-language (L2) learners are limited by the paucity of object descriptions of accented speech. Increases in perceived accent are known to negatively impact judgments of employability (Carlson & McHenry, 2006) and adequacy (Gluszek & Dovidio, 2010). Medical environments, in particular, present a context within which non-native pronunciation may lead to serious communication breakdowns, especially when listeners are elderly (Burda et al., 2009; Cameron & Williams, 1997). Improved outcomes in pronunciation assessment and modification strategies will bolster the achievement of non-native English speakers' social and professional goals in the United States and facilitate communicative effectiveness with Americans who serve as students, colleagues, and patrons to non-native speakers of English. Pronunciation, while intuitively a simple function of one's ability to perceive and produce native (L1) speech sounds with native-like quality, results from a complex interaction among multiple linguistic levels that reflects structural and functional differences between the L1 and L2 (J. E. Flege, 1981; Chen, 1999). The structural and functional differences between Mandarin and American English (AE) have predicted some of the known characteristics of Mandarin Accented English (MAE) and provide a basis for generating hypotheses regarding other factors that influence the extent of perceived accent and the optimal foci for pronunciation modification.

While there is some controversy over the exact size and constituency of the Mandarin vowel system Chen, 1999), the AE vowel system contains as many as twice the number of vowels resulting in a more densely populated vowel space. At least five Mandarin vowels appear to be phonetically similar to ones found in AE (the vowels in beat, bait, boot, boat, and bot). Yet other vowels in AE have no equivalents in Mandarin (e.g., the vowels in the words bit, bet, bat, bought, and hood). Normative values for vowel formant frequencies and segment durations in upper-midwestern AE are well established (J. Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952). Individual vowel durations in AE are either long or short, reflecting the linguistic contrast of tense/lax that does not exist in Mandarin. Vowel production in MAE appears to reflect the confluence of these L1-L2 differences with speakers showing particular difficulties producing accurate distinctions for closely-spaced, tense-lax contrasting vowels (e.g., vowels in the words "beat" versus "bit") (J. E. Flege, Bohn, & Jang, 1997). Additionally, MAE may be characterized by a smaller vowel space than AE with shorter, and less variable vowel durations (Chen, 1999; Chen, Robb, Gilbert, & Lerman, 2001). Across speakers, the amount of similarity to native speakers in the magnitude of the vowel duration contrast effects correlates with perceived foreign accent (J. E. Flege, 1993). While one might predict that MAE vowels common to both L1 and L2 may be most native-like, even for common vowels, MAE vowels appear to differ acoustically from AE vowels (Chen, 1999; Chen et al., 2001). Additionally, there is evidence that front vowels in MAE are acoustically more native-like than back vowels (Olagbaju, Barkana, & Gupta, 2010).

The goal of this study is to provide preliminary comparisons of the perceptual, acoustic, and kinematic characteristics of vowel production in AE and MAE. Perceptual analyses of MAE vowel production are derived from phonetic transcripts. The consistency and equivalence of comparable vowels between MAE and AE is analyzed and the acoustic and kinematic characteristics associated with MAE are determined. While these L1-L2 analyses are largely descriptive in nature, it is hypothesized that acoustic and kinematic measures will reveal some correspondence with the form and degree of perceived vowel pronunciation errors. Morever, two specific hypotheses will be addressed: 1) that MAE vowels are characterized by relatively smaller working space than AE vowels; and 2) that MAE back vowels are particularly distinct from native AE back vowels.

# METHODS

## Database

Data for the current work were derived from a new database of matched acoustic and electromagnetic articulography (EMA) data. The Marquette University EMA-MAE Corpus will be released publicly to help advance research in areas such as pronunciation modeling, acoustic-articulatory inversion, L1-L2 comparisons, pronunciation error detection, and accent modification training. The corpus includes an L1 group of 10 male and 10 female native English (upper-midwestern American English) language speakers and an L2 group of 10 male and 10 female native Mandarin (two dialect regions, including Beijing accent and Shanghai accent Modern Standard Mandarin) Chinese speakers. Articulatory data were collected on a Northern Digital, Inc. Wave Speech Research System with five degree of freedom (5 DOF) sensors (three dimensional position plus two dimensional sensor plane orientation-pitch and roll) at a 400 Hz sampling rate with synchronous acoustic records.

Articulatory sensors in the database include the jaw, lower lip, upper lip, tongue body, and tongue tip placed in the midsagittal plane. In addition, there are two lateral sensors, one at the right corner of the mouth and one in the right central midpoint of the tongue body. A bite plate reference plane calibration allows for the articulatory space to be defined relative to the individual speaker's midsagittal and maxillary occlusal planes, and provides jaw and dental reference information. In addition, a palatal calibration for each speaker provides information about the shape of the hard palate.

The data set includes approximately 45 minutes of data for each speaker, including word, sentence, and paragraph level speech samples. Text-prompted materials include Rogers (1997) list of minimally contrasting words with emphasis on the likely acoustic-phonetic confusions characteristic of MAE, a set of words and pseudo words covering the phonetic space of English vowels (J. Hillenbrand et al., 1995; Peterson & Barney, 1952), sentences from the TIMIT database (Garofolo et al., 1993) and Harvard Intelligibility Sentences (IEEE Subcommittee, 1969), sentences containing words with contrastive stress, and several paragraphs emphasizing aspects of speech such as intelligibility, breath group utilization, accented-English intelligibility, speaking rate, and segmental timing (T. H. Crystal & House, 1982; T. H. Crystal & House, 1990; T. H. Crystal & House, 1988; Green, Beukelman, & Ball, 2004; Honorof, McCullough, & Somerville, 2000; Patel et al., 2012).

## Current Data

### *Participants*

The current data were derived from a subset of ten male speakers from the Marquette University EMA-MAE Corpus. L1 (AE) and L2 (MAE) groups were each comprised of five participants. All participants met the inclusionary criteria of being between the ages of 18-40, being native L1 or L2 speakers, and having no history of speech, language, or hearing pathology, no history of orofacial surgery (other than typical dental extractions), and no history of use of anticonvulsant, antipsychotic, or anti-anxiety medications (as these factors may affect motor performance). Participants read and signed informed consent documents. All procedures and documentation were approved by the Institutional Review Board of Marquette University. Each participant received a $50 cash incentive for participating in the approximately two hour data acquisition for the complete corpus.

### *Procedures*

EMA sensors were adhered to the articulators using Periacryl™ adhesive (Glustich Inc., Delta, BC, Canada). To improve the duration and reliability of sensor adhesion, lingual sensors were bonded with small squares of silk between the sensor and lingual surfaces. Similarly, labial sensors had 2 mm diameter circles of Super Poligrip Strips® denture wax (GlaxoSmithKline Consumer Healthcare L.P.) and the dental sensor had a 3 x 5 mm strip of Stomahesive® periostomal barrier (ConvaTec, Skillman, NJ, USA) used as intermediaries to support adhesion.

Participants were seated in a non-metallic chair (constructed primarily of polyvinyl chloride) with the field generator for the NDI Wave EMA system positioned approximately 5 cm from the head in left profile to assure that all sensors fell within the $300\text{mm}^3$ field setting of the EMA system. Berry (2011) has demonstrated that this setup tends to limit sensor position tracking errors to less than 0.5 mm during dynamic articulator movements. Participants wore a pair of plastic glasses frames with a 6DOF reference sensor attached at midline to allow for use of the EMA system's automated head movement correction algorithm. A computer screen was positioned approximately 1 m away to display all stimuli. Data were collected in a sound-attenuating acoustic booth, with time-synced acoustic records obtained using a cardioid pattern directional condenser microphone (AKG® CS-1000; AKG Acoustics GmbH, Vienna, Austria) positioned approximately 1 m from the center of the electromagnetic field. This microphone positioning was useful in minimizing microphone and EMA system interference.

Data for the current work were obtained following a period of accommodation to the EMA sensors, during which participants spoke casually with the experimenter and lab staff and read a paragraph of text to become accustomed to the size and location of the text-prompted stimuli. Participants read 330 text-prompted words in single-word citation (Rogers, 1997; J. Hillenbrand et al., 1995). Words were blocked into approximately 25 words per data record. This blocking presentation allowed the experimenter to monitor EMA sensor adhesion intermittently and allowed participants to reposition themselves for comfort without making extraneous noise during recordings. The first set of words took the /hVd/ form (J. Hillenbrand et al., 1995) and were produced twice by each participant (once spontaneously and once in imitation of the experimenter). This procedure was used because these were the only stimuli in the corpus that were not necessarily "real" words and pilot data suggested that both L1 and L2 participants demonstrated some difficultly achieving the targeted vowels when reading these words.

### *Vowel Data*

Vowels were analyzed in 166 words spoken by each participant. The vowels /i, ɪ, e, æ, u, ʊ, o, ɑ/ were analyzed in all CV and CVC form words in the database. Citation words with consonant clusters and multisyllabic forms were not included in the current analysis because these linguistic forms do not occur in Mandarin and were assumed to have a non-negligible impact on the measures of interest. Data presented in the current work are derived from three levels of analysis: 1) phonetic transcription; 2) first (F1) and second (F2) formant frequency measures and; 3) spatial positions (x,y) values of the tongue blade (TB) sensor (positioned midsagittally approximately 1 cm posterior to the tongue tip) and the tongue dorsum (TD) sensor (position midsagittally approximately 2 cm posterior to TB).

Primarily broad phonetic transcriptions (diacritic markers were used sparingly) were completed by a graduate student in Speech-Language Pathology who had reasonable prior experience transcribing MAE and had the same dialect base as the AE controls. Transcriptions were analyzed for the occurrence of both vowel and consonant errors to gain estimates of L2 error patterns and intelligibility. While L1 participant productions were not phonetically transcribed, these recordings were assessed perceptually to assure an absence of spoken errors.

Acoustic analysis and vowel boundary indexing were completed using the TF32 software (Milenkovic, 2005). Formant tracks were obtained using pitch-synchronous LPC (26 coefficients). All spectrograms were analyzed with a 300 Hz bandwidth, with a very low noise floor setting, and a frequency range around 5 kHz. F1 and F2 values were manually corrected within visualized windows of approximately 500 ms centered around each spoken word if notable tracking errors were identified by the experimenter. Vowel onset and offset boundaries were defined at the middle of the first and last glottal striations to extend clearly through F1 and F2 as observed with the aforementioned spectrographic settings. The temporal midpoint of the vowel was estimated from the defined vowel boundaries and the nearest four glottal pulses immediately flanking this midpoint were averaged to estimate the F1 and F2 values for each spoken vowel. Vowel space areas were determined using a "vowel multilateral" approach (Fox & Jacewicz, 2008). For any given vowel space calculation, pertinent data were circumscribed by a convex hull. The area within was then determined using the general formula for the area within an irregular, convex polygon:

$$A = \frac{(x_1 y_2 - y_1 x_2) + (x_2 y_3 - y_2 x_3) ... + (x_n y_1 - y_n x_1)}{2} \tag{1}$$

where $(x_1, y_1), (x_2, y_2) ... (x_n, y_n)$ are the *n* vertices of the convex hull.

Kinematic data were re-expressed for each participant within a local, anatomically-defined coordinate system using wax dental biteplate method. Based on a biteplate record, data were transformed to a coordinate space with origin defined at the tips of central maxillary incisors and with axes oriented to align with each participant's maxillary occlusal and midsagittal planes. The resulting coordinate space reflects the anterior-posterior and inferior-superior dimensions of the oral cavity working space within the second quadrant of a Cartesian space (Westbury, 1994). Relative increases in the anterior and superior position of an EMA sensor are described by positive increases in the values of the *x* and *y* coordinates, respectively. Vowel-related EMA sensor positions were determined using the acoustically-defined temporal midpoints described above. For each vowel, the nearest four sensor position samples flanking the midpoint were averaged to estimate the position of the articulatory fleshpoint.

# RESULTS

## Perceptual Analysis

Figure 1 summarizes the results of a perceptual analysis based on broad phonetic transcription. Percent correct "words" (black bar) indicates the proportion of words that were perceived to contain no vowel or consonant errors. The values obtained for the five MAE participants were 57%, 57%, 52%, 45%, and 54% (in ascending order by participant number), consistent with relatively comparable levels of closed-set, word level intelligibility. All other bars in Figure 1 indicate the proportion of vowel replicates transcribed indistinguishably from the AE productions.

These data appear somewhat inconsistent across participants, suggesting some idiosyncrasy in vowel-specific proficiency. Table 1 shows perceived vowel accuracy averaged across participants. Back vowels are perceived to be slightly less accurate than front vowels, with vowel accuracy particularly diminished for low-back vowels.

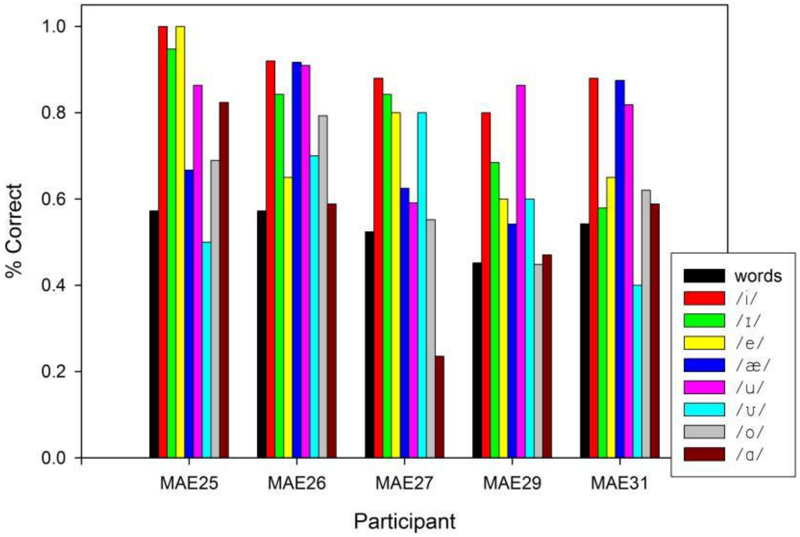

**FIGURE 1.** Summary of perceptual judgments of production accuracy based on phonetic transcriptions of 166 words spoken by MAE participants. Percent correct "words" indicates the proportion of words with broad phonetic transcriptions identical to AE participants (no perceived vowel and/or consonant errors). Percent correct values for each vowel indicate the proportion of words in which the vowel was perceived to be produced without error.

**TABLE 1.** Perceived vowel accuracy averaged across participant.

| | Back Vowels | Front Vowels |
|---|---|---|
| **High Vowels** | /u/ 81% | /i/ 90% |
| | /ʊ/ 60% | /ɪ/ 78% |
| | /o/ 62% | /e/ 74% |
| **Low Vowels** | /ɑ/ 54% | /æ/ 73% |
| (Average) | (64.25%) | (78.75%) |

## Acoustic Analysis

Figure 2 shows acoustic (F1-F2) data from all vowel replicates and all participants coded by group. AE vowels are indicated by yellow data points and MAE vowels are indicated by green data points. Convex hulls were generated for each participant to circumscribe their acoustic working space and are shown in blue (AE) and red (MAE). Table 2 shows calculated vowel space areas for all participants. The largest three vowel spaces were obtained for participants in the AE group and the smallest two vowel spaces were obtained for participants in the MAE group. Group identity is not consistently differentiated by vowel space, though averages across participants within group indicate that the vowel space of the MAE group is modestly smaller than the AE group.

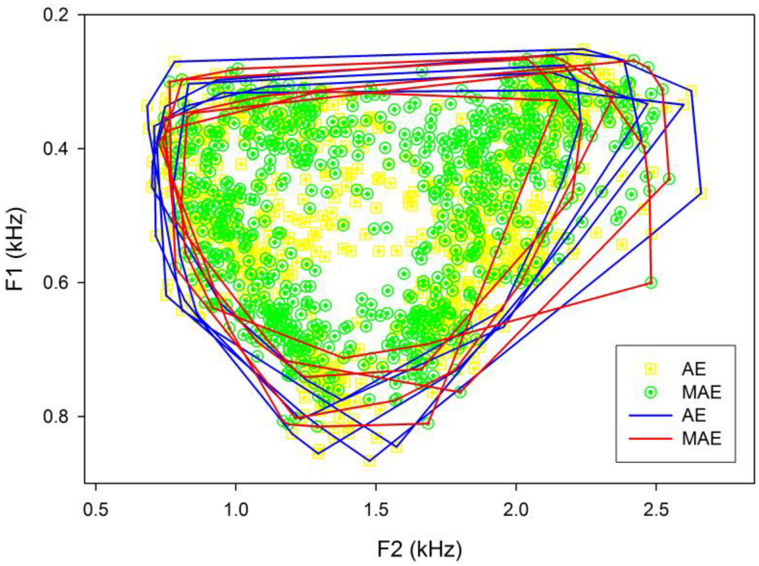

**FIGURE 2.** Acoustic (F1-F2) data from all vowel replicates and all participants with convex hulls used to characterize vowel space. Data are coded by participant group.

**TABLE 2.** Vowel space - area within a convex hull circumscribing each participant's acoustic data (kHz$^2$).

| | **AE Participants** | **MAE Participants** |
|---|---|---|
| | 0.647 (AE32) | 0.624 (MAE25) |
| | 0.556 (AE33) | 0.571 (MAE26) |
| | 0.796 (AE35) | 0.529 (MAE27) |
| | 0.706 (AE38) | 0.487 (MAE29) |
| | 0.569 (AE39) | 0.618 (MAE31) |
| (Average) | (0.655) | (0.566) |

Table 3 shows vowel space calculations completed for individual vowels and averaged across participants within group. These measures are intended to characterize group differences in vowel-specific acoustic variability. The difference column indicates the group differences within vowel. The consistently positive values in this column indicate that for all vowel contexts MAE participants produced vowels (on average) that utilized greater acoustic working space. The two largest vowel-specific differences between groups are the back vowels /u/ and /o/. The two smallest vowel-specific differences between groups occur for the vowels /æ/ and /ʊ/. These observations may be noteworthy since the vowels /u/ and /o/ occur in both Mandarin and English, while the vowels /æ/ and /ʊ/ occur only in English. As a consequence, while we might expect greater variability for unfamiliar (L2 only) vowels, acoustic variability appears to be greater for two vowels that are common to both languages.

**TABLE 3.** Vowel space calculated within vowel averaged across participants with group differences (kHz$^2$).

| **Vowel** | **AE Participants** | **MAE Participants** | **Difference** |
|---|---|---|---|
| /i/ | 0.027 | 0.049 | 0.022 |
| /ɪ/ | 0.043 | 0.083 | 0.040 |
| /e/ | 0.071 | 0.118 | 0.047 |
| /æ/ | 0.046 | 0.060 | 0.014 |
| /u/ | 0.061 | 0.167 | 0.106 |
| /ʊ/ | 0.033 | 0.053 | 0.020 |
| /o/ | 0.106 | 0.208 | 0.102 |
| /ɑ/ | 0.025 | 0.062 | 0.037 |

## Kinematic Analysis

Figure 3 shows kinematic data from all corner vowel replicates for all participants coded by group. High vowels /i/ and /u/ are indicated by yellow squares for AE and green circles for MAE. Low vowels /ɑ/ and /æ/ are indicated by blue, upward-pointing triangles for AE and red, downward-pointing triangles for MAE. The two data clouds shown reflect data from two different lingual fleshpoints. Positions for front vowels were obtained from the TB sensor (approximately 1 cm posterior to the tongue tip). Positions for back vowels were obtained from the TD sensor (approximately 2 cm posterior to the TB sensor along the tongue surface). Apparent overlap in contrasting colors may be indicative of less distinctive fleshpoint positions between high-low vowel contrasts. In this regard, overlap between red and green data points appears more substantial than overlap between blue and yellow data points. This observation is consistent with relatively reduced articulatory contrast along the high-low vowel dimension and seems particularly apparent for the MAE back vowels. Additionally, fleshpoint positions for all MAE vowels appear somewhat higher (larger y) than those for AE vowels and MAE front vowels may be positioned somewhat posteriorly (smaller x) to the AE front vowels.

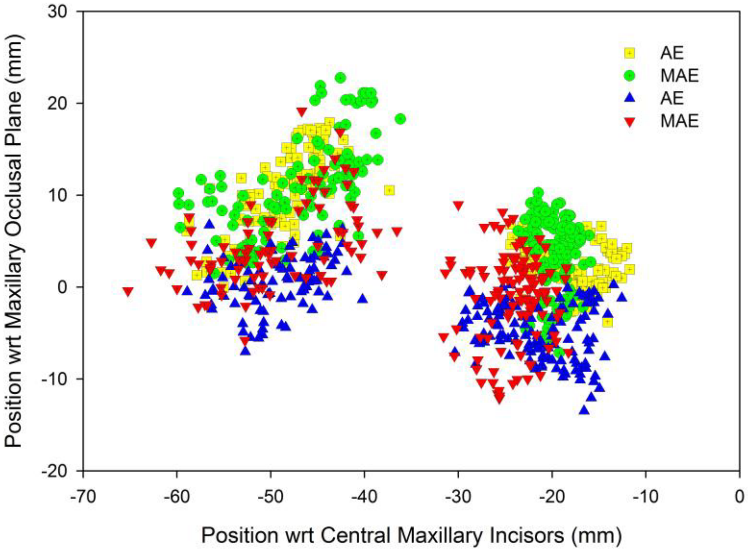

**FIGURE 3.** EMA sensor positions for the corner vowels for all participants. AE participants are shown in yellow (high vowels) and blue (low vowels). MAE participants are shown in green (high vowels) and red (low vowels). The two clouds of points were obtained from the tongue blade sensor (front vowels) and tongue dorsum sensor (back vowels).

**TABLE 3.** EMA sensor positions for corner vowels averaged (SD) within group (mm).

| Vowel | AE-x | AE-y | MAE-x | MAE-y |
|---|---|---|---|---|
| /i/ | -17.91(3.38) | 1.82(2.27) | -19.76(1.67) | 3.55(4.05) |
| /æ/ | -21.29(4.43) | 0.01(2.51) | -24.12(3.18) | 2.54(4.33) |
| /u/ | -48.92(4.01) | 10.14(4.41) | -47.13(5.75) | 10.80(5.30) |
| /ɑ/ | -49.24(4.35) | 0.52(3.05) | -49.81(6.31) | 4.56(4.56) |

Table 3 shows average EMA sensor positions and standard deviations for corner vowels within group. These data are generally consistent with the qualitative observations made above. The spatial distinctiveness of fleshpoint positions for high-low vowel contrasts is reflected in y-value differences. For front vowels, AE participants have an average high-low contrast distance of 1.81 mm (/i/$_y$ –/æ/$_y$ ), while MAE participants have a distance of 1.01 mm. For back vowels, AE participants have an average high-low contrast distance of 9.62 mm (/u/$_y$ –/ɑ/$_y$ ), while MAE participants have a distance of 6.24 mm. This relatively smaller distance between the articulatory positions for different MAE vowels is coupled with somewhat greater position variability (larger SD), particularly for back vowels. Moreover, these data are consistent with the observation made from Figure 3 that all MAE vowels are produced with higher articulatory positions on average and front vowels are produced with more posterior articulatory positions. Taken together, these data are consistent with the interpretation that MAE talkers produce vowels in a smaller articulatory working space using higher and more posterior tongue positions combined with a tendency toward relatively larger vowel-specific variability compared to AE talkers.

## DISCUSSION

While the current results are essentially descriptive and reflect preliminary analyses of data from a limited number of participants, the findings nonetheless offer a useful extension of the limited existing data on MAE vowel production. Chen et al., (2001; Chen, 1999) analyzed F1and F2 measures of 11 vowels from 40 MAE and 40 AE talkers and found significant differences in both familiar and unfamiliar vowels as well as relatively reduced overall vowel spaces for MAE talkers. Their acoustic data are consistent with the prediction that MAE tongue positions tend to assume a more posterior ("backed") position compared to AE talkers. Our findings are consistent with these expectations. Acoustic vowel space tends to be somewhat reduced for MAE talkers compared to AE talkers. Our kinematic data are consistent with the expectation that the tongue is positioned more posteriorly in MAE speech than in AE speech. Additionally, our data suggest that tongue positions are higher on average for MAE talkers, resulting in a compression of the high-low vowel contrasts that is particularly evident for back vowels. Furthermore, this compression of the overall vowel space appears to be combined with relatively larger vowel-specific ranges of variation, suggesting greater overlap among MAE vowel targets compared to AE.

Perceptual data from the current work appear to be consistent with acoustic and articulatory findings. Specifically, the five MAE talkers produced vowels that decreased somewhat in perceived accuracy from high to low and front to back positions. Thus, low vowels were judged to be less consistent than high vowels and back vowels were judged to be less consistent than low vowels. MAE participants varied between 45-57% in the number of words produced without perceived errors. Vowel errors made a non-negligible contribution, suggesting that the characteristics of MAE vowel production contribute to reduced speech intelligibility.

While the finding of a reduced overall vowel space was expected based on the literature that has previously addressed this issue, the finding of increased vowel-specific working spaces for MAE talkers appears to be novel. An important consideration in the current work is that vowel measures were taken from CV and CVC words containing a variety of consonants, some of which do not occur in Mandarin. Thus, compared to previously existing data that are limited to the /hVd/ context, the current work also embraces the impact of diverse phonetic environments on MAE vowel production. Effects of phonetic context on vowel formants are significant (J. M. Hillenbrand, Clark, & Nearey, 2001) and time-varying patterns of formant change may be critical to speech perception (Strange, Jenkins, & Johnson, 1983). The finding of relatively increased vowel-specific working spaces may belie differences in coarticulatory behavior that would not be apparent in phonetically neutral environments such as /hVd/. Since Mandarin is populated by numerous diphthongs and triphthongs (double and triple vowel sequences) while AE has only five, this apparent variability in monophthong production may reflect a MAE tendency toward more "dynamic" vowel productions that is motivated by primary learning of Mandarin. Given these considerations a critical extension of the current work is to examine the characteristics of the time-varying acoustic and kinematic patterns. Such an extension of the current work, when coupled with an analysis of the complete set of talkers from the Marquette University EMA-MAE Corpus, will allow for a more in-depth examination that can be reasonably subjected to statistical testing to help further differentiate MAE and AE vowels.

## CONCLUSIONS

Perceptual, acoustic, and kinematic analysis of vowel production for five AE talkers and five MAE talkers are consistent with the expectation that the MAE vowel space is reduced relative to AE. This reduction is likely a reflection of more posterior and more superior tongue positions for MAE talkers and is coupled with relatively increased variability in vowel-specific working space in MAE compared to AE. Perceptual data suggest that vowel errors make a non-negligible contribution to reduced speech intelligibility. Important extensions of the current work include the examination of more complete acoustic and kinematic event patterns and the examination of a larger pool of participants to allow for robust statistical assessment.

## ACKNOWLEDGMENTS

This work was supported by grant IIS-1141826 from the National Science Foundation.

## REFERENCES

Berry, J. J. (**2011**). Accuracy of the NDI wave speech research system. *Journal of Speech, Language and Hearing Research, **54***(5), 1295.

Burda, A. N., Bradley-Potter, M., Dralle, J., Murphy, J., Ries, S., & Roehs, A. (**2009**). Influence of age and native language on immediate verbal repetition. *Perceptual and Motor Skills, **109***(1), 169-176.

Cameron, R., & Williams, J. (**1997**). Sentence to ten cents: A case study of relevance and communicative success in nonnative-native speaker interactions in a medical setting. *Applied Linguistics, **18***(4), 415-445.

Carlson, H., & McHenry, M. (**2006**). Effect of accent and dialect on employability. *Journal of Employment Counseling, 43*(2), 70-83.

Chen, Y. (**1999**). Acoustic characteristics of american english produced by native speakers of mandarin. (Doctor of Philosophy, University of Connecticut).

Chen, Y., Robb, M., Gilbert, H., & Lerman, J. (**2001**). Vowel production by mandarin speakers of english. *Clinical Linguistics & Phonetics, **15***(6), 427-440.

Crystal, T. H., & House, A. S. (**1982**). Segmental durations in connected speech signals: Preliminary results. *The Journal of the Acoustical Society of America, **72***, 705.

Crystal, T. H., & House, A. S. (**1990**). Articulation rate and the duration of syllables and stress groups in connected speech. *The Journal of the Acoustical Society of America, **88***, 101.

Crystal, T. H., & House, A. S. (**1988**). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America, **83***(4), 1553-1573.

Flege, J. E. (**1981**). The phonological basis of foreign accent: A hypothesis*. *Tesol Quarterly, **15***(4), 443-455.

Flege, J. E., Bohn, O. S., & Jang, S. (**1997**). Effects of experience on non-native speakers' production and perception of english vowels. *Journal of Phonetics, **25***(4), 437-470.

Flege, J. E. (**1993**). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America, **93***(3), 1589-1608.

Fox, R. A., & Jacewicz, E. (**2008**). Analysis of total vowel space areas in three regional dialects of american english. *Journal of the Acoustical Society of America, **123***(5), 3068.

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., Dahlgren, N., & Zue, V. (**1993**). *TIMIT acoustic-phonetic continuous speech corpus* National Technical Information Service.

Gluszek, A., & Dovidio, J. F. (**2010**). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review, **14***(2), 214-237.

Green, J. R., Beukelman, D. R., & Ball, L. J. (**2004**). Algorithmic estimation of pauses in extended speech samples of dysarthric and typical speech. *Journal of Medical Speech-Language Pathology, **12***(4), 149.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (**1995**). Acoustic characteristics of american english vowels. *The Journal of the Acoustical Society of America, **97***, 3099.

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (**2001**). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America, **109***, 748.

Honorof, D., McCullough, J., & Somerville, B. (**2000**). Comma gets a cure: A diagnostic passage for accent study.

IEEE Subcommittee. (**1969**). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, **17***(3), 225-246.

Milenkovic, P. (**2005**). *TF32* - Computer Program (Madison, WI).

Olagbaju, Y., Barkana, B. D., & Gupta, N. (**2010**). English vowel production by native mandarin and hindi speakers. *Information Technology: New Generations (ITNG), 2010 Seventh International Conference on,* 343-347.

Patel, R., Connaghan, K., Franco, D., Edsall, E., Forgit, D., Olsen, L., Ramage, L., Tyler, E., & Russell, S. (**2012**). The caterpillar: A novel reading passage for assessment of motor speech disorders. *American Journal of Speech-Language Pathology* [Epub ahead of print].

Peterson, G. E., & Barney, H. L. (**1952**). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, **24***(2), 175-184.

Rogers, C. (**1997**). Intelligibility of chinese-accented english (doctoral dissertation, indiana university, 1997). *Dissertation Abstracts International, **58***(9)

Strange, W., Jenkins, J. J., & Johnson, T. L. (**1983**). Dynamic specification of coarticulated vowels. *The Journal of the Acoustical Society of America, **74***, 695.

Westbury, J. R. (**1994**). On coordinate systems and the representation of articulatory movements. *The Journal of the Acoustical Society of America, **95***, 2271.