Predictors of L2 Vowel Learning Success Under Biofeedback Training

Joanne Jingwen Li[1], Samantha Ayala[1], Douglas M. Shiller[2], Tara McAllister[1]

[1]New York University; [2]Université de Montréal
jl6132@nyu.edu; saa524@nyu.edu; douglas.shiller@umontreal.ca; tkm214@nyu.edu

ABSTRACT

Previous studies have reported that the visual presentation of acoustic or tongue-kinematic information during speech (i.e., biofeedback) can facilitate speech sound learning; however, participant-level success has been highly heterogeneous, and individual differences have rarely been investigated. The current study explored predictors of success in an L2 vowel learning task under different biofeedback conditions: visual-acoustic or ultrasound. Progress in production accuracy was predicted to be correlated with individual learners' sensory acuity in auditory and somatosensory domains. Baseline accuracy, production variability and phonological awareness were also examined and found to be significant predictors of learning. Sensory acuity was not significant as a main effect or interaction, which might reflect limitations of the acuity measures adopted for this initial study. Future studies will explore additional measures of sensory acuity and will continue to investigate the potential benefit of customizing training paradigms based on individual learners' sensorimotor characteristics.

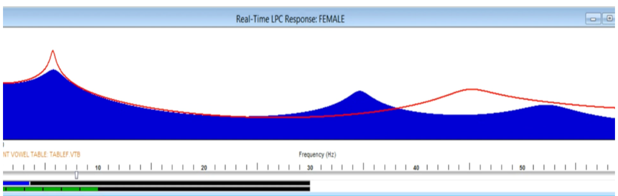
Keywords: L2; biofeedback; individual differences; sensory acuity

1. INTRODUCTION

Visual biofeedback can be used to provide a real-time visual display of speech, together with a model representing target production, to augment the process of speech production training. Various aspects of speech can be presented with different technologies for visual biofeedback, such as place and degree of tongue-palate contact using electropalatography (EPG) [9], tongue shape and movements with ultrasound [1], and spectral properties of the acoustic signal of speech using *visual-acoustic biofeedback* [14]. The current study focuses on ultrasound and visual-acoustic biofeedback.

Previous studies have reported positive effects of ultrasound and visual-acoustic biofeedback training in individuals with speech sound disorders [1, 14, 19] and typically developing speakers acquiring non-native sounds [5, 11]. However, most of these studies have also shown the degree of learning outcomes to be highly heterogeneous across participants [15, 20]. The factors underlying this variability have rarely been investigated. The *personalized learning* framework suggests that individual differences can act as significant predictors of learning success under different training paradigms, and that learning outcomes might be optimized if learners are assigned to a training paradigm aligned with their individual profile of abilities [17, 22]. Inspired by this framework, the current study examined predictors of learning success when participants received ultrasound or visual-acoustic biofeedback (Fig. 1) in a task of learning vowels in a second language (L2), Mandarin.

Figure 1: Visual acoustic biofeedback shows real-time formant frequencies (blue wave) and a superimposed target formant pattern (red line) in a real-time LPC spectrum (generated with KayPentax Computerized Speech Lab/CSL, Model 4500).



The primary predictors under investigation were individual learners' auditory and somatosensory acuity, motivated in part by the DIVA model [6, 7, 8] which suggests that speakers' precision in production is influenced by the size of target regions in auditory and somatosensory space. In addition, individual speakers' reliance on these two sensory domains has been shown to be independent of one another, with some speakers showing a sensory preference for one over the other [4, 13, 16]. Therefore, we predicted that there would be subsets of typical speakers who have relatively higher acuity in one domain and poorer acuity in the other. Further, we hypothesized that different profiles of sensory acuity could interact with biofeedback type to predict learning outcomes. When learning to produce non-native speech sounds, speakers who have relatively poorer auditory acuity may have difficulty establishing accurate acoustic targets from auditory input. These speakers would be expected to benefit the most from visual-acoustic biofeedback, since the visually presented acoustic targets could compensate for the speakers' less precise auditory perception. On the other hand, speakers who have intact auditory but poor somatosensory acuity may perceive an L2 target accurately but may not be able to use tactile-kinesthetic information to identify the articulator movements needed to generate that target. These speakers would be expected to benefit more from ultrasound biofeedback, since information about articulator placement that is typically received through somatosensory channels is made visually explicit through this type of biofeedback, which could enhance the establishment of somatosensory targets.

Based on previous findings, baseline production accuracy [11], production variability [10] and phonological awareness (PA) [17] were also examined as potential predictors.

2. METHODS

2.1. PARTICIPANTS

Participants were sixty-five female native speakers of English (18-30 years old) with no prior knowledge of Mandarin or other languages that involve the target vowels described below. All participants reported normal hearing and speech language abilities and gave informed consent before the experiment. The study received approval from the Institutional Review Board at New York University.

2.2. STIMULI

An L2 vowel learning task was conducted using two Mandarin vowel targets, /y/ and /u/, which mainly differ in the second formant (F2) and tongue position along the front-back dimension. The stimuli for imitation and training were isolated /y/ and /u/ productions from female Mandarin speakers. Additional /y/ and /u/ tokens produced by a female Mandarin speaker were synthesized into a 240-step /y-u/ continuum using STRAIGHT [12]. All recordings (.wav) took place in a sound-attenuated booth, with a sampling rate of 44 kHz and encoding of 16-bit.

2.3. PROCEDURES

Participants completed two sessions on separate days with approximately one week between sessions.

2.3.1. SESSION I

Session I began with a pure-tone hearing screening, followed by tests of auditory acuity, somatosensory acuity and PA.

Auditory acuity was tested using an AXB staircase sound discrimination task. In each trial, participants listened to three sounds in a sequence and were required to indicate if it was the first (A) or the third (B) sound that was different from the second one (X). Sounds were drawn from the /y-u/ continuum described above. Acoustic distance between A and B in each trial was adjusted based on the accuracy of the response in the previous trial: step size decreased following correct responses and increased following incorrect responses.

Oral somatosensory acuity was then tested using a Spatial Resolution Acuity task [21], in which participants were instructed to use their tongue tip to identify capitalized letters of 7 different sizes embossed on one end of Teflon strips. This task also followed a staircase design, in which the first trial started with a medium letter size and accuracy in each trial determined the letter size for the next attempt.

Participants took both acuity tasks twice, but only scores from the first run were used due to evidence of a possible learning effect.

Lastly, PA was tested using the composite score from the elision, blending and phoneme isolation tasks of the Comprehensive Test of Phonological Processing (2^{nd} edition; CTOPP-2).

2.3.2. SESSION II

Participants were randomly assigned to either a visual-acoustic or ultrasound biofeedback condition and received 30 min (20 blocks of 6 trials) of training for each of the two Mandarin vowels. The order of vowels being trained was counterbalanced across participants. During the training, participants repeated after audio models while viewing real-time visual-acoustic or ultrasound feedback. A target image was superimposed on the LPC spectrum or ultrasound screen for participants to match. A female Mandarin native speaker provided qualitative feedback after each block of trials.

Twenty repetitions of each target vowel were audio recorded at baseline and after training of each vowel for acoustical analysis. No feedback was provided for these repetitions.

2.4. ANALYSES

First (F1) and second (F2) formant frequencies from the mid-point of each isolated vowel production were measured and converted to Bark scale for both English participants and Mandarin control speakers. Euclidean distance (ED), which quantifies the distance from English participants' productions to the center of the distribution of productions from their target speaker, was used to evaluate production accuracy. Learning was quantified as the difference in median ED between pre-training and post-training time points, with a negative value indicating improved accuracy. Production variability was quantified as the area of an ellipse [10] representing a 95% confidence interval around the distribution of productions for each individual and vowel.

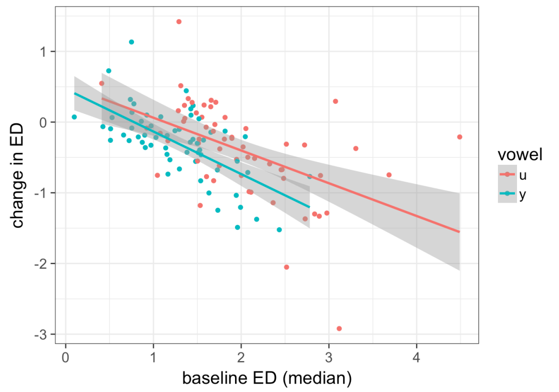
3. RESULTS

Scores for each Session I measure were reasonably dispersed across individual speakers, but the acuity measures were not normally distributed, with auditory acuity showing signs of a ceiling effect. Accordingly, we avoided parametric assumptions in our analyses.

Descriptive statistics showed a general reduction of ED from pre- to post-training phases, indicating improvement in production accuracy over the course of training. Effect sizes (Cohen's d) revealed that the magnitude of change was moderate for both /u/ and /y/. Linear regression models were used to examine which factors were significant in predicting the magnitude of change over the course of biofeedback training. Change in ED was entered as the dependent variable, and independent variables included auditory acuity, somatosensory acuity, PA, baseline production accuracy, production variability at both baseline & post-training time points, and biofeedback type (visual-acoustic or ultrasound). The model also included the theoretically predicted interaction between biofeedback type and the two domains of sensory acuity.

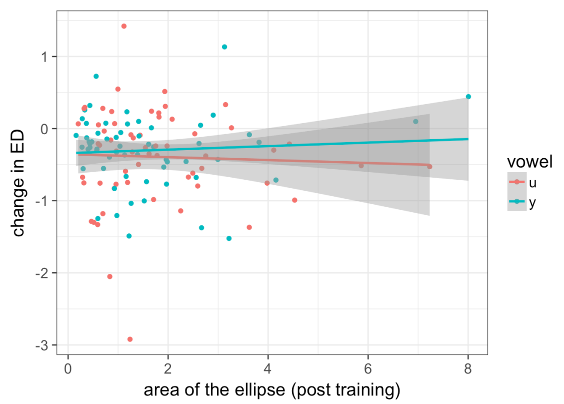
Model results showed that baseline production accuracy was a significant predictor for both /y/ (β = -0.66, SE = 0.1, $p < .001$) and /u/ (β = -0.47, SE = 0.1, $p < .001$), with the negative coefficients indicating that a lower baseline accuracy was associated with greater improvement for both vowels (Fig. 2).

Figure 2: Association between baseline production accuracy (baseline ED) and change in ED over the course of training



For /y/ only there was a significant effect of post-training variability (β = 0.09, SE = 0.037, p = 0.02), in which greater learning was associated with lower variability at the post-training time point (Fig. 3). For /u/ only, PA was a significant predictor (β = -0.02, SE = 0.0072, p = 0.01), showing that higher PA was associated with a greater improvement.

Figure 3: Association between post-training production variability (area of the ellipse) and change in ED over the course of training



The rest of the hypothesized predictors and interactions did not reach significance.

4. DISCUSSION

For both vowels, participants showed a general reduction in ED from pre- to post-training phases, indicating that they did get closer to the L2 targets, despite the brief duration of training provided. This corroborates previous research evidence showing that even a short duration of biofeedback training can yield improved speech production [5, 11]. It is unknown if the observed improvement could be specifically attributed to the biofeedback component. This could be investigated in future studies by including a no-biofeedback control group. The present study also did not find a significant difference in response to the two types of biofeedback examined, although further study is needed to confirm that these approaches are truly equivalent in their effect.

The finding that lower baseline accuracy was associated with greater learning gains is consistent with previous research [2, 11]. This finding makes intuitive sense because low baseline performance leaves more room for improvement and avoids ceiling effects.

The present finding that a larger magnitude of progress was associated with lower post-training variability for /y/ aligns with previous findings [10, 11] suggesting that successful learners show a reduction in variability corresponding with the establishment of a new and stable category for a non-native sound. In contrast, there was no significant effect of post-training variability for /u/. This result contrasts with previous work suggesting that effects of variability are more pronounced for L2 vowels that have a similar counterpart in L1 (as /u/ does) versus dissimilar vowels [10]. However, it is possible that some speakers showed low variability for the similar L2 vowel not because they established a stable L2 target, but because they were simply using the motor plan for English /u/ both before and after training. Given this possibility, variability may be less reliable as a sign of learning success in the context of an L1-adjacent vowel like /u/.

For /u/ only, PA was a significant predictor, with higher PA corresponding with a larger magnitude of progress in production accuracy. Previous research has reported PA as a predictor of L2 learning (e.g., [17]), but the mechanism of its association with L2 production is less well understood. One possibility is that speakers with higher PA are more likely to attend to subtle acoustic differences observed across languages. This could also explain why PA was a significant predictor for the similar vowel /u/ but not the uncategorized vowel /y/, which learners are likely to perceive as distinct even without close attention to phonetic detail.

Contrary to our predictions, there was no significant interaction between biofeedback type and sensory acuity, nor were there main effects of sensory acuity. This might reflect limitations of the current acuity measures. The synthesized /y-u/ continuum used in the AXB discrimination task was generated from the /y/ end, meaning that the task was likely a better measure of acuity perceiving /y/ than /u/. However, in light of the above discussion of the relatively close proximity of Mandarin and English /u/ and its implications for perceptibility, a continuum starting from the /u/ end might be better for the current purpose. It is also possible that production variability, which has been suggested to reflect how narrowly specified a speaker's auditory targets are [3, 18], is simply a better or more relevant measure of auditory acuity than the explicit AXB task used here. Similarly, although stereognosis tasks similar to that used here have been effective as predictors of production distinctness for consonants such as [s] and [ʃ], its focus on tactile acuity of the tongue tip means that it may be less relevant for the vowel targets that constitute the focus of the present study. In addition, some participants achieved a perfect score in the second run, suggesting that a more challenging task (e.g., drawing from a larger set of letters) might be more appropriate.

5. CONCLUSION

The current study explored predictors of learning in biofeedback-enhanced L2 speech sound training. We hypothesized that sensory acuity in auditory and somatosensory domains would interact with biofeedback type (visual-acoustic vs. ultrasound) to predict response to training. The hypothesized interaction was not significant, but the potentially relevant factor of production variability did predict learning outcomes. Phonological awareness was also a significant predictor, consistent with previous findings [17]. Future research in this line could ultimately enhance clinical and pedagogical outcomes by pairing learners with training paradigms customized to their own sensorimotor characteristics.

6. REFERENCES

[1] Adler-Bock, M., Bernhardt, B. M., Gick, B., Bacsfalvi, P. 2007. The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *American Journal of Speech-Language Pathology, 16*(2), 128-139.

[2] Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., Tohkura, Y. i. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J Acoust Soc Am, 101*(4), 2299-2310.

[3] Franken, M. K., Acheson, D. J., McQueen, J. M., Eisner, F., & Hagoort, P. (2017). Individual variability as a window on production-perception interactions in speech motor control. *The Journal of the Acoustical Society of America, 142*(4), 2007-2018.

[4] Fucci, D. 1972. Oral vibrotactile sensation: An evaluation of normal and defective speakers. *Journal of Speech, Language, and Hearing Research, 15*(1), 179-184.

[5] Gick, B., Bernhardt, B., Bacsfalvi, P., Wilson, I., Zampini, M. 2008. Ultrasound imaging applications in second language acquisition. *Phonology and second language acquisition, 36*, 315-328.

[6] Guenther, F. H. 1994. A neural network model of speech acquisition and motor equivalent speech production. *Biological cybernetics, 72*(1), 43-53.

[7] Guenther, F. H. 1995. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological review, 102*(3), 594.

[8] Guenther, F. H. 2016. *Neural control of speech*: Mit Press.

[9] Hardcastle, W. J., Gibbon, F. E., Jones, W. 1991. Visual display of tongue-palate contact: Electropalatography in the assessment and remediation of speech disorders. *British Journal of Disorders of Communication, 26*(1), 41-74.

[10] Kartushina, N., Frauenfelder, U. H. 2014. On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in psychology, 5*, 1246.

[11] Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., Golestani, N. 2015. The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds a. *J Acoust Soc Am, 138*(2), 817-832.

[12] Kawahara, H., Morise, M., Banno, H., Skuk, V. G. 2013. *Temporally variable multi-aspect N-way morphing based on interference-free speech representations.* Paper presented at the Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific.

[13] Lametti, D. R., Nasir, S. M., Ostry, D. J. 2012. Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience, 32*(27), 9351-9358.

[14] McAllister Byun, T., Hitchcock, E. R. 2012. Investigating the use of traditional and spectral biofeedback approaches to intervention for/r/misarticulation. *American Journal of Speech-Language Pathology, 21*(3), 207-221.

[15] McAllister Byun, T., Hitchcock, E. R., Swartz, M. T. 2014. Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research, 57*(6), 2116-2130.

[16] Nasir, S. M., Ostry, D. J. 2008. Speech motor learning in profoundly deaf adults. *Nature neuroscience, 11*(10), 1217-1222.

[17] Perrachione, T. K., Lee, J., Ha, L. Y., Wong, P. C. 2011. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *J Acoust Soc Am, 130*(1), 461-472.

[18] Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America, 116*(4), 2338-2344.

[19] Preston, J. L., Leaman, M. 2014. Ultrasound visual feedback for acquired apraxia of speech: A case report. *Aphasiology, 28*(3), 278-295.

[20] Preston, J. L., McCabe, P., Rivera-Campos, A., Whittle, J. L., Landry, E., Maas, E. 2014. Ultrasound visual feedback treatment and practice variability for residual speech sound errors. *Journal of Speech, Language, and Hearing Research, 57*(6), 2102-2115.

[21] Steele, C. M., Hill, L., Stokely, S., Peladeau-Pigeon, M. 2014. Age and strength influences on lingual tactile acuity. *Journal of texture studies, 45*(4), 317-323.

[22] Wong, P. C., Vuong, L. C., Liu, K. 2017. Personalized learning: From neurogenetics of behaviors to designing optimal language training. *Neuropsychologia, 98*, 192-200.