ACOUSTIC AND ARTICULATORY QUALITIES OF SMILED SPEECH

Megan Keough, Avery Ozburn, Elise Kedersha McClay, Michael David Schwan, Murray Schellenberg, Samuel Akinbo, Bryan Gick*
University of British Columbia

1 Introduction

Studies have shown that listeners can use acoustics to identify smiled speech [1,2,3], which is characterised by increased amplitude, higher f0 [1], and some increases in formants [3,4]. Though smiling changes vocal tract shape [5], little is known about its effect on articulation, and how those articulatory changes affect the acoustics, nor whether results from single-speaker studies such as [4] apply to a larger population. Our study addresses this gap through a multi-speaker production experiment examining the articulation and acoustics of smiled and neutral English vowels, focusing on formant values, lip spreading, lip protrusion, lip angle, and larynx height. Facial positions were tracked following [4], and larynx height was measured with ultrasound following [6]. We hypothesize that smiled speech, compared to neutral speech, is characterised by higher f0, higher formant frequencies, raised larynx, and spread lips with corners turned up.

2 Methodology

Results are from 10 native English speakers (5 male, 5 female). Three target words (key, caw, coo) were used to elicit the vowels /i/, /a/, and /u/ in the carrier sentence 'I got a_toy', and were presented in E-Prime [7] alongside an image of a toy. Stimuli were grouped into 8 blocks of randomized all-neutral or all-smiled sentences, with 3 repetitions per sentence per block; there was also a training block. Neutral and smiled blocks alternated, with first block type randomly assigned. To remind participants of the target facial expression, stimulus slides were followed by images of neutral or smiling children from [8]. 16 dot stickers were placed at anthropometrically-defined points on participants' faces (Fig. 1).

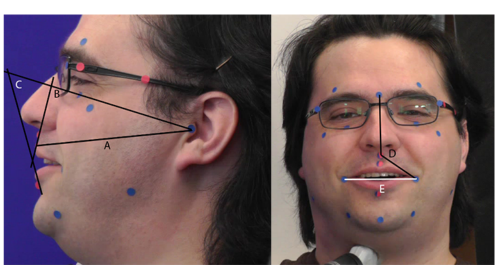

Figure 1: Anthropometric dots. A=Lip Protrusion Length (LPL), B=Upper Protrusion Angle (UPA), C=Lower Protrusion Angle (LPA), D=Lip Corner Angle (LPA), E=Lip Spread Length (LPL)

An Aloka Pro-Sound SSD 5000 ultrasound machine with an Aloka UST-9119-3.5 convex transducer (pulse frequency 3.5MHz, field of view 120º) collected larynx images. Following [6], the probe was positioned manually against the right thyroid lamina near the laryngeal prominence; it was kept stable with a mechanical arm. The audio (digitized at 44.1kHz) was recorded using a Sennheiser MKH 8060 shotgun microphone. Two Panasonic HC-V700 cameras recorded facial video.

Target vowels from the audio were delineated by hand. Praat [9] scripts marked the vowel midpoints, extracting midpoint f0, F1, and F2 values. A JPEG from each vowel midpoint was extracted from all videos. Facial markers were used to measure distances (in pixels) and angles in ImageJ [10]. For ultrasound images, annotators measured the distance from the most stable lamina point to the edge of the frame and converted to centimetres using the pixel length of the ultrasound's 10cm ruler.

3 Results and discussion

Measures were normalized within-speaker using a z-score transform. Data was subset by vowel, and for each (normalized) measure, a mixed effects model was run (using the lme4 package [11] in R [12]) with the measure as a dependent variable, a fixed effect of condition (smiled vs. neutral) and a random effect for subject with a random slope for condition.

3.1 Acoustic results

As Fig. 2 shows, smiling significantly ($|t|\geq 2$) raised f0 for all three vowels (t = 6.13, 6.19, 5.97 for *caw, coo, key*), an expected result based on previous literature [1,3,13].

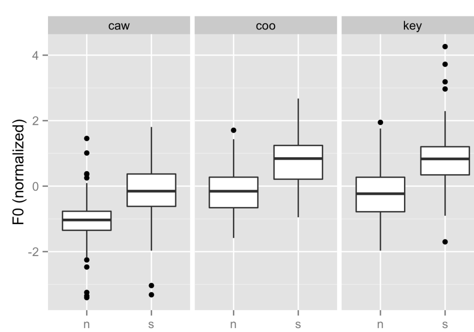

Figure 2: Normalized F0 for each vowel in neutral (n) and smiling (s) conditions.

* gick@mail.ubc.ca

However, while [1] found that all formant frequencies were higher in smiled speech, we found that smiling only significantly raised F1 and F2 for *caw* (t = 3.19, 3.26 respectively); F1 and F2 results for the other vowels were not significant. Our data thus align more with [13], who found that f0 was one of the most important factors in perceiving smiling. Further, we would expect /a/ to be affected most by smiling, since only in /a/ are the lips phonetically unconstrained; in /i/, they are spread, and in /u/, rounded. However, these results do not support findings of increased formants for only /i/ in German [4].

3.2 Larynx Measurements

The results showed no significant effect of smiling on larynx height. This is one way in which our results depart from [13], which found larynx height and f0 to be the main predictors of speech being perceived as smiled. Also, since we saw a significant effect in f0, and since larynx height is a main influence on f0 [14], the lack of larynx height difference in smiling is notable.

3.3 Articulatory Results

Lips were significantly more spread in smiled than in neutral speech (t = 12.64, 4.75, 18.08 for *caw*, *coo*, *key*). LCA was smaller for all three vowels (t = -9.79, -8.13, -7.22 for *caw*, *key*, *coo*), meaning lip corners were raised. This indicates that participants smiled, with lips spread and lip corners turned up. LPL was significantly less for smiled speech for *caw* (t = -2.88) and *key* (t = -5.21), but not for *coo*. Similarly, UPA and LPA were significantly smaller and larger respectively for *caw* and *key* (t = -2.98, 2.28 for *caw*, t = -5.98, t = 3.49), but not significant for *coo*. The lack of protrusion effect for /u/ suggests that lip spreading and rounding do not conflict. These results diverge from [4], which showed a suppression in the rounding of /u/ in German, possibly due to differences between German and English /u/. UPA and LPA measurements are novel to this study; they correspond to the standard LPL, suggesting these angles are informative and potentially useful.

4 Conclusion

The effects of smiling observed here on articulation were lip spreading and lip corner raising, and raised f0 for all three vowels, but raised F1 and F2 only for /a/. Larynx height was not significantly affected, contrary to findings in [13].

Acknowledgments

We thank: Molly Babel, Kyra Borland-Walker, Amon Ge, Thomas J. Heins, Michael McAuliffe, and Nick Romero. This project has been made possible by a NSERC grant to the final author. Errors are our own.

References

[1] Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. Perception & Psychophysics, 27 (1), 2427. doi:10.3758/Bf03199901

[2] Quené, H., Semin, G., & Foroni, F. (2012). Audible smiles and frowns affect speech comprehension. *Speech Communication*, 54 (7), 917922. doi:10.1016/j.specom.2012.03.004

[3] Torre, I. (2014). Production and perception of smiling voice. Proceedings of the first Postgraduate and Academic Researchers in Linguistics at York (PARLAY 2013) conference.

[4] Fagel, S. (2010). Effects of smiling on articulation: Lips, larynx, and acoustics. In A. Esposito, N. Campbell, C. Vogel, A. Hussain, & A. Nijholt (Eds.), Development of Multimodal Interfaces: Active Listening and Synchrony Lecture Notes in Computer Science Volume 5967 (294303). Berlin: Springer.

[5] Shor, R. E. (1978). The production and judgment of smile magnitude. *The Journal of General Psychology*, 98(1), 79-96. doi: 10.1080/00221309.1978.9920859.

[6] Moisik, S. R., Lin, H., & Esling, J. H. (2014). A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association*, 44(01), 21-58.

[7] Psychology Software Tools, Inc. (E-Prime 2.0). (2012). Retrieved from http://pstnet.com

[8] Egger, H. L., Pine, D. S., Nelson, E., Leibenluft, E., Ernst, M., Towbin, K. E., & Angold, A. (2011). The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): a new set of children's facial emotion stimuli. *Int. J. Methods Psychiatr. Res*., 20: 145–156. doi: 10.1002/mpr.343

[9] Boersma, P. & Weenink, D. (2015). Praat: doing phonetics by computer. Version 5.4.06, Available from http://praat.org/

[10] Rasband, W. S. (2014) ImageJ. Available from http://imagej.nih.gov/ij/

[11] Bates D., Maechler M., Bolker B., and Walker S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7, http://CRAN.R-project.org/package=lme4.

[12] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Satistical Computing, Vienna, Austria. *http://www.R-project.org*. 2008.

[13] Lasarcyk, E. & Trouvain, J. (2008). Spread lips + raised larynx + higher f0 = smiled speech? - An articulatory synthesis approach. *Proceedings of the 8th International Seminar on Speech Production*, Strasbourg, France, 8-12 December 2008, pp. 345- 348.

[14] Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in FO control. *Language and Speech*, 42(4), 401.