INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Acoustic and Articulatory Study of Ewe Vowels: A Comparative Study of Male and Female

Kowovi Comivi Alowonou[1], *Jianguo Wei*[1], *Wenhuan Lu*[1], *Zhicheng Liu*[1], *Kiyoshi Honda*[1], *Jianwu Dang*[2]

[1]College of Intelligence and Computing Tianjin University, Tianjin, China
[2]Japan Advanced Institute of Science and Technology, Ishikawa, Japan
kalowonou@tju.edu.cn, jianguo@tju.edu.cn, wenhuan@tju.edu.cn, fsliuzhicheng@tju.edu.cn

Abstract

In order to investigate the difference in Ewe males and Ewe females during the production of Ewe vowels, results from the comparative quantitative and qualitative assessments of tongue shape and movement using ultrasound imaging as well as the comparative evaluation of F1 and F2 frequency values from data collected from 9 Ewe male speakers and 6 Ewe female speakers, were presented in this study. The results showed that vowels are produced with higher formant frequencies by Ewe female speakers compared to Ewe male speakers, except for the vowel /ɛ/ produced with a lower F1 frequency by Ewe females. The articulatory results showed a higher and more forwarder tongue configuration for Ewe male compared to female counterparts.

Index Terms: acoustic analysis, articulatory analysis, ewe vowel, ssanova, ultrasound

1. Introduction

The sounds of human language are the result of the airflow from the lungs through the vocal tract. In other word, in human language, words and sounds are produced by a combination of the lips, tongue, jaw, teeth, palate and larynx, which modulates vocal-tract resonance and airflow from the lungs. Therefore, acoustic and articulatory information is essential when investigating speech production of languages.

The languages of developing countries received less attention in the past [1, 2], and as consequences, their physiological mechanism is almost unknown. Acoustic and articulatory investigation of these languages would provide not only a good understanding of their speech production mechanism but also good information to facilitate the implementation of speech processing services such as automatic speech recognition. Thus, in this study, we built a multimodal speech database containing acoustic data and articulatory data though ultrasound experiments on of Ewe language, and conducted an investigation on Ewe oral vowels, by a quantitative and qualitative analysis of the tongue configuration, from the ultrasound images; as well as an acoustic analysis.

This work is organized as follow: In section 2, we presented the methodology for data collection, analysis, and measurement. Section 3 presents our findings and, discussions, as well as the conclusion, are given in section 4.

2. Method

2.1. Speakers and Stimuli

There were nine males and six females, between the ages of 22 and 34, chosen for this study (median age 25.8). The speakers are students and workers grew up in Togo with no reported history of neurological disorders or diseases, or any speech, language or hearing difficulties.

The word list contained CV sequences where C was balanced for the place of articulation using the voiceless stop consonants /p t k/ and V Ewe oral vowels (/a/, /e/, /ɛ/, /i/, /o/, /ɔ/, /u/). Those segments comprised 3 * 7 = 21 tokens with forms like /ka/, /pi/, and /tu/. The speakers produced 63 individual utterances, i.e., 3 repetitions of each vowel.

2.2. Instrumentation and Recording Procedure

Data collection were made by using a Telemed Echo Blaster 128 CEXT-1Z [3], an audio interface Roland Octa-Capture UA-1010 and an Articulate Instruments pulse-stretch unit [[4], all connected via USB to a laptop running Windows software. Besides, we use a BEHRINGER ECM8000 condenser microphone and an Articulate Instruments stabilization headset [3]. Acoustic signals and ultrasound images were recorded simultaneously with AAA software (Articulate Assistant Advanced) [5] and synchronized with the Articulate Instruments pulse-stretch. The ultrasound images were acquired with a frame a frame rate of 95 fps, using a 5–8 MHz convex probe with a probe curvature radius of 10 mm, set to 5 MHz, a depth of 90 mm and a field of view of 90%. The audio signal was captured at a sample rate of 22.05 kHz with a 16-bits resolution and saved as WAV files.

The recording took place in a soundproof recording studio, where each participant was recorded separately. They were asked to produce each word three times, displayed on the laptop screen placed in front of them. The Articulate Instruments Stabilization headset was used to fix the ultrasound transducer at 90º, to the submandibular region of the speakers, in order to visualize the tongue in the sagittal plane and to avoid unwanted movements during the recording period.

http://dx.doi.org/10.21437/Interspeech.2019-2196

2.3. Acoustic and Articulatory Measurements

2.3.1. Acoustic Measurements

Before the analysis, the acoustic files were first exported from the AAA software in WAV format and were down-sampled to 11.025 kHz and pre-emphasized (98%).

Additionally, the acoustic files were labeled with the PRAAT software [6]. Furthermore, a script was written for the PRAAT software using linear predictive coding algorithm in order to estimate the center frequencies for the first two formants (F1, F2). The parameters for formant analysis were set as: the number of formants 5, max formant 5000 Hz for male speakers and 5500 Hz for female speakers, and dynamic range 30 dB. The vowels were Hanning windowed (25ms) with an overlap of 50%. The formant values are shown in Table 1.

To deal with the unwanted anatomical/physiological talker-specific variation that can affect acoustic measurements in this study, F1 and F2 frequencies were normalized using Lobanov z-score transformation [7]. The results, however, are not in Hertz-like values. Therefore, normalized formant frequencies were rescaled into Hertz-like values [8].

Table 1: *Mean F1 and F2 frequency from Ewe male and female speakers.*

Vowels	**F1 (Hz)**		**F2 (Hz)**	
	F	M	F	M
/a/	750	703	1450	1347
/e/	468	397	1967	1955
/ɛ/	537	555	1810	1751
/i/	352	294	2157	2107
/o/	485	415	1237	806
/ɔ/	617	585	1397	1032
/u/	397	341	1450	781

2.3.2. Articulatory measurements

Using the Articulate Assistant Advanced (AAA) software, tongue contours were tracked automatically across the entire utterance; however, some manual adjustment was applied whenever it was necessary. The tongue spline data points were exported at the temporal midpoint of the vowels into the Cartesian coordinates for plotting purpose using SSANOVA [9], as described in section 2.4. Since the tongue surface typically approximates an arc more closely than it approximates a horizontal line, the tongue spline data points were also exported into the polar coordinates with the angular coordinate (Θ) and radial coordinate (r). The goal behind this is to use the radius as input for the LME model described in section 2.4. The input data therefore assume a circle for the tongue shape. The palate was obtained by recording the participants swallowing three times. The frame used to obtain the spline for the palate was the frame obtained when the tongue dorsum was visibly pressed against the palate during the swallow at the most anterior and superior position, highlighting the alveolar ridge. The palate location was identified manually and drawn on the ultrasound images using the AAA software.

2.4. Statistical Analysis

2.4.1. Smoothing Spline ANOVA (SSANOVA)

The SSANOVA is a statistical method for comparing curves, which is often used in phonetics for comparison of the entire tongue contours that are obtained from ultrasound or MRI [9]. The Cartesian coordinates extracted from the tongue spline data points were submitted to a SSANOVA model to plot the best-fit curve across repetitions. These best-fit curves can be interpreted as the tongue shape for a particular vowel's production. It must be noted that the SSANOVA is used in this study for the plotting purpose only. Thus, for the SSANOVA plots, the zero on the y-axis denotes the probe origin, and the zero on the x-axis denotes the probe centre.

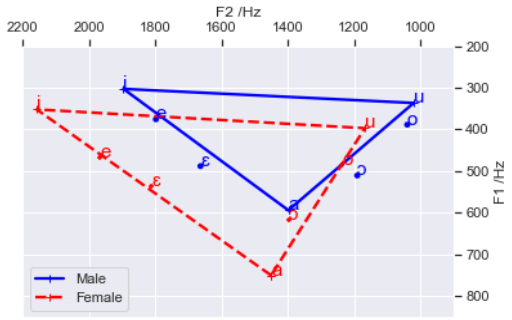

Figure 1: *Ewe female acoustic vowel space (dashed) and Ewe male acoustic vowel space (solid).*

2.4.2. Linear Mixed Effects Analysis (LME)

We used R and lme4 package, to perform an LME analysis of the relationship between formant frequencies and gender [10, 11] in order to estimate the difference in formant frequencies between the vowels produced by male and female speakers. As fixed effects we entered gender. Speaker and context were used as random effects. The formant frequencies (F1/F2) were used as the dependent variables. To obtain the P-values we used a likelihood ratio test, which compares the likelihood ratio of the full model with the effect in question (gender) against the likelihood of the model without the effect in question [12].

Regarding the quantitative measurement of the tongue, an LME analysis was conducted to estimate the distance of the tongue front/back location from the origin of the coordinate system. Doing so we can quantify how far the tongue moved front/back during the vowels production, while a comparison is made between male and female speakers. The radius of the polar coordinates mentioned earlier was used as the dependent variable, gender as a fixed effect, and speaker and context as random effects. The difference in the distance of the back/front of the tongue between male and female speakers is then obtained. If the distance of the front of the tongue from the origin is superior for male than for female during the production of a vowel, it means that the vowel is produced with the higher tongue body by male compared to female. Likewise, if the distance of the back of the tongue from the origin point is superior for male compared to female, when produced a vowel, it means that the tongue is more retracted for male than for female. These differences are given in millimeters. It has to be precise that the same model was for each vowel for the statistical analysis.

Table 2: *LME results for gender difference in F1/F2 for the seven oral vowels (model F1/F2 ~ gender). (f = female, m = male).*

Linear Hypotheses	Estimate (Hz)		Std Err		t-value	
	F1	F2	F1	F2	F1	F2
/a_m/-/a_f/=0	-83.4	-117.3	19	36.30	-11.72	-10.8
/e_m/-/e_f/=0	-65.74	-642.5	18.2	69.8	-2.264	-4.906
/ε_m/-/ε_f/=0	31	-193.6	37.97	34.6	1.2	-8.166
/i_m/-/i_f/=0	-58.4	-55.3	17.7	18.3	-3.1	-2.4
/o_m/-/o_f/=0	-92.58	-986.09	24.41	62.85	-9	-36.30
/$ɔ_m$/-/$ɔ_f$/=0	-47.6	-621.76	17.3	57.63	-2.53	-11.8
/u_m/-/u_f/ =0	-51	-187.1	18.1	32.4	-2.1	-25

Table 3: *LME results for gender difference in back/front position of the tongue for the seven oral vowels (back/front ~ gender). (f = female, m = male).*

Linear Hypotheses	Estimate (mm)		Std Err		t-value	
	back	front	back	front	back	front
/a_m/-/a_f/=0	-0.9965	5.284	2.2629	3.468	-18.302	1.524
/e_m/-/e_f/=0	-1.693	0.8442	3.131	1.0640	-0.541	0.793
/ε_m/-/ε_f/=0	-1.671	1.327	2.851	1.894	-0.586	0.701
/i_m/-/i_f/=0	5.994	-0.5771	2.985	1.1212	2.008	-0.515
/o_m/-/o_f/=0	-0.07454	5.154	2.71662	3.910	-0.027	1.318
/$ɔ_m$/-/$ɔ_f$/=0	-2.030	4.423	1.794	3.704	-1.131	1032
/u_m/-/u_f/ =0	-0.2043	3.338	3.2959	4.101	-0.062	781

3. Results

3.1. Acoustic Results

The acoustic vowel space of male and female speakers is plotted in Figure 1 (male in blue/solid and female in red/dashed). From Figure 1, we observed that in the vertical dimension, the female acoustic space moves downward compared to the male acoustic space. In the horizontal dimension, the female acoustic space moved forward compared to the male acoustic space.

Regarding the results presented in Table 2, we have observed that concerning the front vowels /i/, /e/ and /a/, F1 frequency values measured from Ewe female speakers are higher than for those from Ewe male speakers. The difference is estimated to 58.4 Hz, 65.74 Hz, 83.4 Hz respectively. The likelihood ratio test reveals that the difference is significant (X2(1) = 5.2717, p = 0.03061; X2(1) = 29.421, p = 0.04398; X2(1) = 7.213, p = 5.444e-07). However, we observed a lower F1 frequency for vowel /ɛ/ when produced by Ewe female speakers. The difference was found not significant (X2(1) = 0.3764, p = 0.4623) and was estimated to 31 Hz. The difference in the F2 frequency domain was also observed. However, unlike that in the F1 frequency domain, the values of F2 are greater for all the front vowels produced by Ewe female speakers compared to the male counterparts. The difference is significant (X2(1) = 6.3123, p = 0.02043; X2(1) = 29.906, p = 2.8e-09; X2(1) = 29.776, p = 2.848e-08) for all the front vowels concerned, by about 55.3 Hz, 642.5 Hz and 117.3 Hz, respectively.

Regarding now to the back-oral vowels /u/, /ɔ/ and /o/, we notice that the F1 and the F2 frequencies measured from the female speakers are both higher than those measured from Ewe male speakers. The likelihood ratio test showed that the difference is significant (X2(1) = 6.121, p = 0.01785; X2(1) = 5.3163, p = 0.02105; X2(1) = 32.835, p = 1.596e-05) and was estimated to 51 Hz, 47.6 Hz and 92.58 Hz respectively.

3.2. Articulatory Results

From the estimated tongue curves plotted in Figure 2, we observed certain differences in the tongue configuration during the production of vowels by Ewe female and male speakers. The output plots of the SSANOVA indicate that the vowels /a/, /e/, /ɛ/, /o/, /ɔ/ and /u/ demonstrate higher and more retracted tongue position when produced by Ewe male speakers, unlike the data from Ewe female speakers where the vowels are produced with lower and more retracted tongue positions. As for the vowel /i/, the tongue is higher in the vertical dimension and more retracted in the horizontal dimension for Ewe female speakers compared to Ewe male speakers.

Considering the front vowels, the estimation of the difference in horizontal displacement of the tongue between the female and male speakers, given by the LME model, is about 0.9965 mm for the /a/ and 1.693 mm, 1.671 mm, 5.994 mm respectively for the vowel /e/, /ɛ/ and /i/. However, this difference is only significant with the vowel /i/ (X2(1) = 3.112, p = 0.07772). Regarding the vertical displacement of the tongue, there is no significant difference in the tongue shape when compared between the female and male speakers, but it was estimated to 5.284 mm, 0.8442 mm, 1.327 mm and 0.5771 mm respectively for the vowel /a/, /e/, /ɛ/ and /i/.

For the back vowels, we observed the same difference in tongue configuration when a comparison is made between Ewe female speakers and Ewe male speakers. The vowel /o/, /ɔ/ and /u/ are produced with a higher and more retracted tongue configuration by Ewe male speakers compared to Ewe

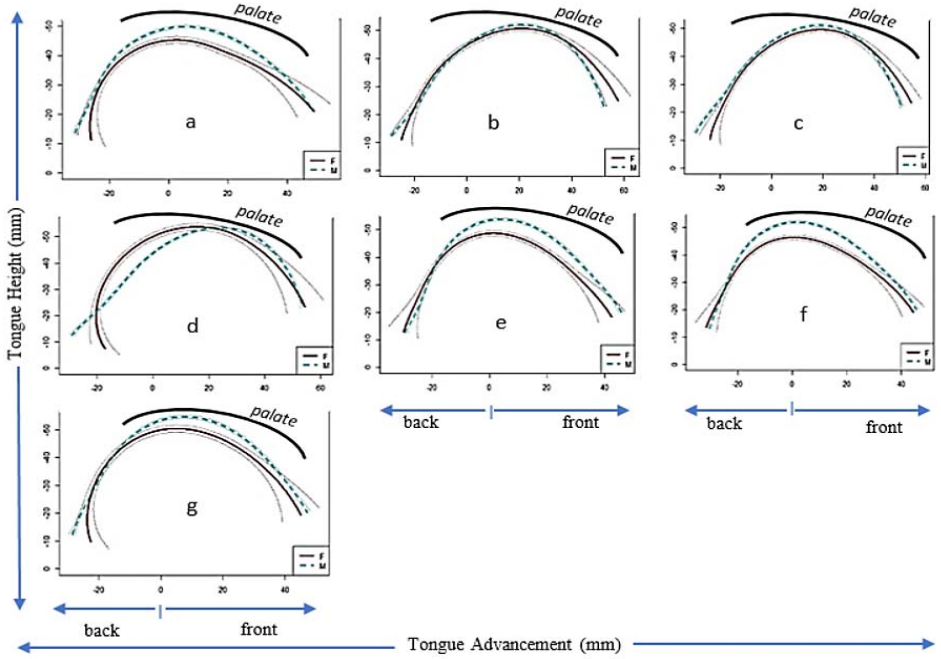

Figure 2: *SSANOVA plot of tongue contours. Gender difference in tongue contours when produced oral vowels (blue-dashed=male, red-solid=female). a: /af-am/; b: /ef-em/; c: /ɛf-ɛm/; d: /if-im/; e: /of-om/; f: /ɔf-ɔm/; g: /uf-um/. (f = female, m = male).*

female speakers. The difference, however, is not significant, and was estimated to 5.154 mm for /o/, 4.423 mm for /ɔ/ and 3.338 mm for /u/, regarding to the vertical dimension, and to 0.07454 mm, 2.030 mm and 0.2043 mm for /o/, /ɔ/, /u/ respectively regarding to the horizontal dimension.

4. Discussions and Conclusion

The formant frequency analysis pointed out cross-gender differences. The vowel formant frequencies appeared to be generally higher for Ewe female speakers, except for the vowel /ɛ/, realized with a lower formant frequency compared to Ewe male speakers. Furthermore, we also observed from Figure 1 that the female vowel space seemed larger, being located lower and more forward compared to the male vowel space. This observations are aligned with the argument of Diehl in [13] for American English and Whiteside in [14] for British English. Based on the argument given by Fant [15] regarding the difference in the physiological of the vocal tract between male and female, we conjecture that these differences in the formant frequencies and in the acoustic vowel space are linked to the fact that male speakers have larger laryngeal cavities and a proportionally longer pharynx than female speakers [16]. Concerning the vowel /ɛ/ produced with a lower F1 frequency by Ewe females, we can suggest that constriction in the lower pharynx of female speakers can affect F1 to be lower compared to male speakers [17]. This result also can be explained by the effect of lip rounding and/or protrusion on F1 frequency as it is known to decrease all formants [18].

The results of the articulatory study revealed that the tongue is lower for our female speakers and more forwarder, compared to the male speakers. We have found that there is a mapping between the articulatory configuration predicted by the acoustic results and the articulatory results obtained from the lingual gesture examination. This observation agrees with Johnson [17] and Stevens [18] who suggested that tongue raising is correlated with a decrease in F1, while tongue lowering is correlated with an increase in F1. Tongue retraction is correlated with a decrease in F2, while tongue advancement is correlated with an increase in F2. However, we found that there is not clear acoustic-articulatory mapping, regarding the vowel /i/ and /ɛ/. We suggest that the formant frequencies can be affected by another articulator.

In summary, we have conducted an acoustic and articulatory investigation on Ewe vowels produced by male and female speakers in order to explore the differences in acoustic and articulatory measures between gender in the Ewe language. We found differences indeed, in both acoustic and tongue configurations. This study completes the work of NADA GBEGBLE [19] by establishing descriptions of articulatory configuration of Ewe vowels, which has never been subjected to investigation before. The results obtained from our study are also useful since it provides good information necessary for the implementation of speech processing services.

5. Acknowledgements

This work is supported in part by grants from the National Natural Science Foundation of China (General Program No. 61471259, and No. 61573254) and in part by NSFC of Tianjin (No. 16JCZDJC35400).

6. References

[1] Besacier, L., et al., *Automatic speech recognition for under-resourced languages: A survey.* 2014. **56**: p. 85-100.

[2] Le, V.-B., L.J.I.T.o.A. Besacier, Speech,, and L. Processing, *Automatic speech recognition for under-resourced languages: application to Vietnamese language.* 2009. **17**(8): p. 1471-1482.

[3] Scobbie, J.M., A.A. Wrench, and M. van der Linden. *Head-Probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement*. in *Proceedings of the 8th International seminar on speech production*. 2008.

[4] Wei, J., et al., *Multi-modal recording and modeling of vocal tract movements.* 2016. **75**(9): p. 5247-5263.

[5] Instruments, A.J.E., UK: Articulate Instruments Ltd, *Ultrasound Stabilisation Headset User's Manual, Revision 1.3*. 2008.

[6] Boersma, P.J.G.i., *Praat, a system for doing phonetics by computer.* 2002. **5**.

[7] Thomas, E.R. and T. Kendall, *NORM: The vowel normalization and plotting suite.* 2007.

[8] Thomas, E.R., et al. *Two things sociolinguists should know: Software packages for vowel normalization, and accessing linguistic atlas data.* in *Workshop at New Ways of Analyzing Variation (NWAV)*. 2007.

[9] Mielke, J.J.T.J.o.t.A.S.o.A., *An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons.* 2015. **137**(5): p. 2858-2869.

[10] Pinheiro, P.J.h.c.r.-p.o.w.p.n., *Linear and nonlinear mixed effects models. R package version 3.1-97.* 2010.

[11] Winter, B.J.a.p.a., *Linear models and linear mixed effects models in R with linguistic applications.* 2013.

[12] Winter, B.J.a.p.a., *A very basic tutorial for performing linear mixed effects analyses.* 2013.

[13] Whitfield, J.A. and A.M.J.I.j.o.s.-l.p. Goberman, *Articulatory-acoustic vowel space: Associations between acoustic and perceptual measures of clear speech.* 2017. **19**(2): p. 184-194.

[14] Whiteside, S.P.J.J.o.t.I.P.A., *Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences.* 1996. **26**(1): p. 23-40.

[15] Fant, G.J.S.T.L.Q.P. and S. Report, *A note on vocal tract size factors and non-uniform F-pattern scalings.* 1966. **1**: p. 22-30.

[16] Chiba, T. and M. Kajiyama, *The vowel: Its nature and structure*. 1941: Tokyo-Kaiseikan.

[17] Johnson, K., *Acoustic and auditory phonetics*. 2011: John Wiley & Sons.

[18] Stevens, K.N., *Acoustic phonetics*. Vol. 30. 2000: MIT press.

[19] Gbegble, N., *Spectrographic analysis of Ewe vowels.*