

SEX DIFFERENCES IN VOCALIC DURATION PRODUCTION IN L1 AND IN L2

Amanda Post da Silveira, Judith Varkevisser
Federal University of São Carlos, Leiden University
psyphon.ap@gmail.com, j.m.varkevisser@biology.leidenuniv.nl

ABSTRACT

This study intends to investigate the duration of American English vowels as produced by males and by females who were first language (L1) and second language (L2) speakers of the language. Results showed that females produced longer vowels than males both in L1 and in L2 speech, which was solely an effect of sex, while no interaction of speech nativeness was found. When comparing our results to studies investigating languages other than English, we observed that similar patterns of sex-related nativeness-independent vocalic duration differences emerged. Finally, we argue for sex differences to be always considered in phonetic studies so that important phenomena may not be overlooked.

Keywords: Speech production, Vocalic duration, Nativeness, Sex-specific phenomena

1. INTRODUCTION

It is well-established in the phonetic literature that males and females (which we will represent here with the Sex variable) present morphological differences or even sociolinguistic features that make them produce speech differently from each other e.g. [1, 2]. Average pitch is higher in females than in males, also the vowel formants have higher center frequencies in females than in males. Less has been said in the literature about inter-sex differences in vowel duration production.

Studies have recurrently shown that female speakers of German, Swedish and American English languages as L1 produce longer vowels than male speakers [3, 2, 4]. Those studies used read sentences as well as spontaneous speech. Results support a biomechanical account to explain the male-female difference in vocalic durations. The explanation goes towards a temporal lengthening of vowels by females in order to compensate for a shorter vocal tract than males in order to achieve similar phonetic targets in terms of vocalic production. The difference in duration is achieved by moving their tongues in different speeds, females move their tongues slower than males. Evidence for the biomechanical hypothesis comes from the studies above mentioned, which show consistent cross-linguistic similarities in

their results. These results have important implications for sex dynamics in speech synthesis and typology [4].

Both in [4] and [2] another possible explanation is given that not only greater vocalic durations may be produced by females, but also greater duration differences between categories as a consequence of sociophonetic factors related to the female sex. For instance, women produce clearer speech than men, because her vocalic production may be more stable concerning spectra and durations, which includes a greater contrast between the durations of long and reduced vowels.

The purpose of the present study was to investigate how male and female speakers of American English as L1 and as L2 produce full and reduced vowels of American English. We conducted an English word naming study. The resulting production data allowed us to check if females and males differ in their vowel durations and if there was symmetry in vowel duration cross-linguistically within the same sex - e.g. females produced similar vocalic duration patterns both in AE L1 and AE L2.

Because vowel reduction in the L1 sound inventory of Brazilian Portuguese L1 (BP L1) speakers is not phonemic, it is most likely that the word naming data of BP L1 speakers should show a contrast relative to English speakers, e.g., reduced vowels in unstressed position might be produced as full vowels.

2. METHOD

2.1. Participants

For this study, 20 subjects took part in the experiment. They were exchange students at Leiden University Centre for Linguistics (LUCL), Leiden, The Netherlands. The American English monolingual group consisted of 10 native monolingual speakers of American English (5 males and 5 females, age range between 19 and 48, mean age: 25.6 years). The non-native American English speaker group consisted of 10 Brazilian participants with Brazilian Portuguese as their first language and English as their second language (5 males and 5 females, ages ranged between 21 and 44 years, mean age: 30). L2 speakers of American English reported their scores in English

proficiency tests, such as TOEFL and IELTS proficiency tests. As only three participants had performed the IELTS test, we transformed their scores into TOEFL scores, using the metrics provided by the TOEFL comparison tool [5]. After the transformation of scores, it was possible to select participants who graded equal to 90 or above according to TOEFL scores (Mean TOEFL Scores = 94.9, SD = 7.4). The L2 participants also performed the X_Lex2.05 English vocabulary test [6]. In order to be selected for the experiment, a participant should score above 3,500 points in the vocabulary test. Their scores ranged between 3,500 and 4,850 points (Mean=4,193 points, SD = 418) out of a maximum of 5,000 points. The TOEFL and the X_Lex2.05 scores rated them as upper intermediate to advanced learners of English. None of the participants reported having hearing problems and all had normal or corrected-to-normal vision. Participation in the experiment and travel costs were paid. Some volunteered to participate with no compensation.

2.2. Stimulus Materials

We used 360 3-syllable words, half of them with stress on the first syllable and the other half with stress on the second syllable, such as *Advocate* and *Adjective*, with stressed full vowels in the first syllable (capitals indicate stressed syllables), as opposed to *gaLACtic* and *goRilla* with unstressed reduced vowels in the initial syllable (underline indicates unstressed syllable).

2.3. Procedure

The experiment took place at LUCL Phonetic lab. Participants were positioned in a sound-attenuating booth and seated at approximately 60 cm from a computer screen (size: 32.5 x 24 cm). The word naming experiment was implemented and presented with E-Prime 2.0 software. English words were presented in black Arial font with size 18 points, at the center of the screen. The list of English stimulus words was randomized for each participant and divided in four experimental blocks containing 90 words each which were counterbalanced in the number of stimuli with stressed and unstressed target syllables. Before each block there was a break. Words were preceded by a 250 ms fixation cross (+) and presented individually for 1000 ms. Participants had another 3000 ms after the word disappeared to pronounce it before another trial began. They were instructed to read the word on screen aloud as fast and as accurately as possible. A Sennheiser MKH-416 unidirectional condenser microphone was used to capture participants' speech directly onto a PC (22,050 Hz, 16 bit/sample) and E-Prime 2.0 recorded

and stored their speech productions for later analysis. Sessions lasted 30 minutes on average. The actual experiment was preceded by a training phase with the same dynamics of the actual experiment and contained 20 American English words not included in the experimental phase.

2.4. Phonetic transcriptions

A Praat [7] script was used to automatically trace the speech responses within the 3-seconds response time window. Word boundaries were later corrected manually by the authors of this study. A second Praat script was used to label the words. Next, a third Praat script was used to automatically align and label the segments according to canonical transcriptions for each word. The automatic alignment and labelling were later checked and corrected by two phonetician experts. Ambiguous transcriptions were discarded by the reviewers and 9 words that resulted in many ambiguities in transcription were excluded from our analyses.

2.5. Vowel duration measurement

Concomitantly to vowel spectrum, vowel duration is an important feature in the relationship of vowel phonemes with word and sentence stress in English. We therefore measured the vowel durations of words in our dataset. In Brazilian Portuguese (the first language of the AE L2 group of this study), only spectrally full vowels are found, whether in stressed or unstressed syllable positions [8]. Also, vowel duration differences between stressed and unstressed vowels are not large in BP (except for variation found in spontaneous speech, see [9]); however, in English the ratio between full and reduced vowels is larger than in BP [10]. We therefore expect that BP speakers of English will show a smaller duration ratio for reduced vs. full English vowels than native English speakers. Note that in our English target words, full vowels only occurred in stressed position, while reduced vowels only occurred in unstressed position.

3. RESULTS

Each participant produced 360 trisyllabic words of which 180 had stress on the first syllable and 180 on the second syllable, making a total of 7,200 word tokens.

Only vowels in the first and second syllables of the word tokens were included in our analyses to avoid effects of final lengthening. Thus, 14,040 vowel tokens, and after data cleaning a total of 11,459 vowel tokens were used for the statistical analysis.

Table 1 shows the results for vowel duration per language group, as well as the ratio of full stressed

vowels vs. reduced unstressed vowels produced on the first and second syllables of trisyllabic English words.

Table 1: Vowel duration means by Language Group and Sex

	Full vowels		Reduced vowels		Ratio full/reduced	
	males	fem.	male	fem.	males	fem.
AE L1	102 (34)	116 (43)	59 (28)	71 (36)	1.73	1.63
AE L2	107 (39)	116 (39)	74 (36)	82 (40)	1.45	1.41

*Vowel duration means in milliseconds and Standard Deviations within parenthesis.

Table 1 indicates that native speakers of American English produced a greater difference in duration between full and reduced vowels in English than Brazilians did. The trend by both males and females from the AE L2 group seems to be to produce a smaller ratio between full and reduced vowels than their peers from the AE L1 group, which is shown by the ratio values.

Concerning Sex differences in the duration of vowels, we observed that females appear to have produced longer vowels than males in both full vowel category and in reduced vowel category. Although the AE L1 group is formed by different participants from the AE L2 group, we noticed the same pattern emerging in the vowel duration of males and females: females produce longer vowels than males both in their L1 and in the L2. The means in Table 1 indicate that the vowel durations of females from the AE L2 group is closer to same-sex AE L1 group than in the two groups of male speakers.

Although male and female differences in vocalic duration are not great (Table 1), in between 8-11 ms for reduced vowels and 9-14ms for full vowels, they are within the range of Just-Noticeable Difference (JND) of vocalic duration. As shown in several papers, JND of vocalic duration is roughly 10% of the total vowel duration e.g. [11, 12, 13, 14, 15], especially in the elicited speech modality and high-quality recorded data, like the data from this experiment. We expect that JND of vocalic duration to vary depending on speech modality (whispered speech and spontaneous speech, for instance), nonetheless, we hypothesize the vocalic duration JND to be symmetrical to our current findings and to be perceived in communication.

A Multiple Regression Analysis with fixed factors was performed in R [16] to investigate how Language Group (American English as L1 or L2), Sex (male or female speakers), and stress (full or

reduced vowel durations) affect vowel duration in American English as produced by L1 and L2 speakers (see Table 2).

Table 2: Multiple Regression Analysis on duration by Language Group, Vowel Quantity and Sex

	Estimate	SE	t	p
(Intercept)	-0.30	.11	-2.63	.08
Language Group	0.22	.07	3.07	.02
Vowel Quantity	-1.16	.16	-7.23	<.01
Sex	0.43	.07	6.02	<.01
Vowel Quantity*Sex	-0.06	.10	-0.58	.57
Language Group*Vowel Quantity	0.22	.10	2.16	.03
Language Group*Sex	-0.10	.05	-2.28	.02
Language Group*Vowel Quantity*Sex	0.01	.06	0.17	.87

We observed main effects of Language Group, Sex, and Vowel Quantity, as well as significant interactions between Language and Sex, and Language and Vowel Quantity. These results support the conclusions drawn from Table 1: The significant main effect of Vowel Quantity indicates that full and reduced vowels in English differ in duration. The duration of vowels in general was also clearly different in the two language groups, being longer in the L2 group than in the L1 group –due to the fact that the L2 durations for unstressed vowels were too long relative to the L1 norm.

The observed interaction between Vowel Quality and Language Group indicates that overall differences in the production of full and reduced vowels depend on whether English vowels were produced by Americans or Brazilians. Furthermore, both males and females showed significant differences in English vowel duration depending on whether they were English L1 or English L2 speakers (as reflected in the interaction between Language Group and Vowel Quantity).

The evidence indicates that full vowels and reduced vowels are produced distinctly depending on sex and language group. To find out whether Brazilians and Americans showed a similar duration reduction to English vowels, a closer look is needed at the reduced vowel durations. The reduced vowels present in this dataset are the mid-central vowel (also known as Schwa, phonetic symbol /ə/), and the near-front near high vowel (phonetic symbol /I/), which

were produced either on the first or second syllables of the trisyllabic words. Overall, 5,673 reduced vowels were available in our dataset. After outliers were excluded, a total of 5,535 items remained for analysis.

The statistical analysis showed that Schwa vowels are produced overall with longer durations than Mid-High Mid-Front vowel vowels. This is due to the intrinsic characteristics of closed and open vowels. Mid vowels, such as Schwas, require longer articulation time due to the jaw opening than high vowels, such as Mid-High Mid-Front vowels.

Males and females from both Language Groups produce comparable differences in vowel durations, which are relatively larger in Brazilians' productions than in Americans'. Male American English speakers produced shorter vowels than female American English speakers in both L1 and L2 groups. To test if the differences between Brazilians' and Americans' vocalic reduction are significantly different, we ran two Multiple Regression Analysis, one for Mid-High Mid-Front vowels and another one for Schwa vowels, taking as a dependent variable the Square root duration values and as independent variables Language Group (American English as L1 or L2) and Sex. In total, 3,600 Schwas were produced. After outliers were excluded, a total of 3,474 tokens remained for statistical analysis.

A Multiple Regression Analysis with Language Group and Sex factors showed that Language Group is significant ($\beta = 0.35$, $SE = 0.08$, $p < .001$), and Sex is significant ($\beta = 0.23$, $SE = 0.08$, $p = .04$) as main factors, but insignificant in interaction ($p > .05$). The production of Mid-High Mid-Front vowel numbered 2,073 tokens, but after outliers were excluded, 2,040 tokens remained for statistical analysis. The results from the Multiple Regression Analysis showed the same pattern as in the Schwa analysis, Language Group and Sex are significant as main factors ($(\beta_{Language\ Group} = 0.29$, $SE = 0.08$, $p < .001$) and $(\beta_{Sex} = 0.31$, $SE = 0.08$, $p < .001$)), but they are not significant in interaction ($p > .05$). As no significant interaction between Language and Sex was found, it indicates that the effects of Language Group and Sex on vowel duration are independent of each other.

4. CONCLUSION

Following studies which investigate sex-specific vowel duration differences, we indicated that female speakers produced longer vowels than male speakers when speaking in their L1, in this case, American English. This pattern was also found in the productions of a non-native speaking group of American English. Our statistical results showed no dependency between Language Group and Sex, so

the inter-sex phenomenon of vowel duration patterns here observed, longer vowels produced by females, is something that occurs independently of the language used by speakers or language of dominance (L1 or L2) and is solely related to sex.

The findings of this study have some important implications for phonetics studies. They provide further evidence for the study of inter-sex phonetic features as independent phenomena. They also show that inter-sex duration differences in vocalic realization are independent of language proficiency because the pattern emerged also in L2 productions.

For future investigations, using data that address specifically male and female speech differences, speech rate is an important factor to be considered for future analyses.

Although this is a preliminary study on the topic, the current data support the hypothesis that females produce longer vowel than males as an articulatory strategy to compensate for vocal tract dimensional differences and not because of a language specific factor, such as language proficiency. These preliminary analyses support differences in the use of language by female speakers and male speakers, as advocated by several studies, e.g. [1, 2, 3, 4, 17, 18, 19] adding to those studies new evidence that female speakers produce larger vocalic duration contrast than male peers, not only in first language, but also in second language.

ACKNOWLEDGEMENTS

I would like to thank Erasmus Mundus and CNPq for funding the PhD project from which this paper originated. I am truly thankful for the dialogue with Vincent van Heuven and Plínio Barbosa about the theme of this paper.

5. REFERENCES

- [1] Simpson, A. P., & Ericsson, C. (2003). Sex-specific duration differences in English and Swedish. In Proceedings of the 15th International Congress of Phonetics Sciences, 1113–1116.
- [2] Simpson, A. P. (2003). Possible articulatory reasons for sex-specific differences in vowel duration. In Proceedings of the 6th International Seminar on Speech Production, Sydney, Australia.
- [3] Simpson, A. P. (1998). Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel AIPUK 33.
- [4] Ericsson, C., & Ericsson, A. M. (2001). Sex differences in vowel duration in read Swedish: preliminary results. Working Papers, 49, 34-37.
- [5] ETS - Measuring the Power of Learning.® Copyright © 2015 by Educational Testing Service.

- [6] Meara, P., & Milton, J. (2006). X-Lex: the Swansea Vocabulary Levels Test. In Coombe, C., Davidson, P. and Lloyd D. (Eds.) *Proceedings of the 7th and 8th Current Trends in English Language testing (CTELT) Conference TESOL Arabia, UAE*, 4, 29-39.
- [7] Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341-347.
- [8] Nobre, M. A., & Ingemann, F. (1987). Oral vowel reduction in Brazilian Portuguese. In R. Channon & L. Shockey (eds.). *In Honour of Ilse Lehiste* (pp. 195-206). Dordrecht, Foris Publications.
- [9] Barbosa, P. A.; Eriksson, A.; Åkesson, J. (2013). Cross-linguistic similarities and differences of lexical stress realisation in Swedish and Brazilian Portuguese. In E.L. Asu, & P. Lippus (Eds.), *Nordic Prosody. Proceedings of the XIth conference, Tartu 2012* (pp. 97-106). Frankfurt am Main: Peter Lang.
- [10] Major, R. (1985). Stress and rhythm in Brazilian Portuguese. *Language*, 61, 259-282.
- [11] Ruhm, H. B., Mencke, E. O., Milburn, B., Cooper, W.A., & Rose, D. E., (1966). Differential Sensitivity to Duration of Acoustic Signals, *Journal of Speech, Language and Hearing Research*, 9, 371-384.
- [12] Huggins, A. W. F. (1972). Just Noticeable Differences for Segment Duration in Natural Speech. *J. Acoust. Soc. Am.* 51, 1270-1278.
- [13] Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual Evidence. *The Journal of the Acoustical Society of America*, 59, 1208-1221.
- [14] Nooteboom, S. G., & Doodeman, G. J. N. (1980). Production and perception of vowel length in spoken sentences. *Journal of the Acoustic Society of America*, 67(1), 278-287.
- [15] van den Broecke, M. P., & van Heuven, V. J. (1983). Effect and artifact in the auditory discrimination of rise and decay time: Speech and nonspeech. *Percept. Psychophys.* 33, 305-313.
- [16] R Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- [17] Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, 3 (2), 621-640.
- [18] Haan, J., & V. J. van Heuven 1999. Male vs. female pitch range in Dutch questions. In *Proceedings of the XIV International Conference of Phonetics Sciences (ICPhS), San Francisco*, 1581-1584. Haas, A. (1979). Male and Female Spoken Language Differences: Stereotypes and Evidence. *Psychological Bulletin*, 86 (3), 615-626.
- [19] Post da Silveira, A. (2016). Word stress in second language word recognition and production. *Enschede: Printkamp*, pp. 270.