HIGH VARIABILITY PRONUNCIATION TRAINING (HVPT): USING DIFFERENT STIMULUS TALKERS AND SOUND-TO-SYMBOL ASSOCIATIONS IN L2 PRONUNCIATION LEARNING

Kelsey Labbé, Brock University

This paper describes a study of High Variability Pronunciation Training (HVPT), investigating the effects of utilizing both different sound symbols and male versus female talkers on L2 perception and production. Using the web application English Accent Coach, eight English language learners were assigned to one of four groups, trained through exposure to either male or female talkers. They were asked to indicate what vowels they heard using either the International Phonetic Alphabet (IPA) or a vowel chart constituted by colours rather than symbols. After hearing each sound stimulus item, learners responded by clicking on an IPA symbol or colour that represented the sound they believed they heard, and received immediate feedback. Descriptive statistics based on pre- and post-tests indicate that participants improved in L2 vowel perception and production, regardless of which sound symbol was used as a reference. In addition, results suggest that the sex of the stimulus talker may have a role in learning for both L2 perception and production. Due to this paper's exploratory nature, future research is needed to best inform practice and to confirm potential impacts of sex of stimulus talker in HVPT and L2 pronunciation learning.

INTRODUCTION

In English language classrooms, pronunciation is a skill that can pose a challenge for teachers, as it is not tied to proficiency and thus requires more individualized instruction; for example, a beginner can have excellent pronunciation whereas an advanced learner can be difficult to understand (Thomson & Derwing, 2015). To address students' unique needs, teachers have often turned to computer software (Breitkreutz et al., 2001). One increasingly popular software innovation is High Variability Pronunciation Training (HVPT). This perception training provides users with numerous auditory stimuli spoken by multiple talkers, and allows the learner to choose which sounds to focus on (Thomson, 2018b). Not only does this type of auditory training improve L2 perception (Lively et al., 1993; Thomson, 2018b) and L2 production (Iverson et al., 2012), but by including talker variability, it can help learners to better generalize sounds to new talkers (Nishi & Kewley-Port, 2007). It remains unclear, however, what type of response symbols HVPT systems should use to represent target sounds. Also, it is not known whether the sex of the stimulus talkers matters. This exploratory study investigates both variables in to determine their potential impact on pronunciation acquisition.

Sound-to-Symbol Association

One important element of HVPT involves exposing learners to multiple auditory stimuli, in which target sounds are associated with a word or symbol; learners must choose which sound they

perceive by choosing from various response options (Thomson 2018b). According to Jones (2017), to help students accurately produce certain vowel sounds, it is useful to introduce a vowel chart, either the Colour Vowel Chart (CVC) or a comparable chart utilizing the International Phonetic Alphabet (IPA). Current research reveals that L2 vowel perception gains are possible using the CVC (Taylor & Thompson, 2012, as cited in Brinton, 2017), but also with other symbols such as nautical flags (Thomson, 2011). Given these gains with different types of symbols, this study aims to confirm whether one type of symbol results in better gains over another, or if the importance is more tied to the training itself.

Talker Quality

One factor not often considered in L2 pronunciation acquisition is talker quality – that is, lower pitch (male) vs. higher pitch (female). Students are often taught by both male and female instructors throughout their language learning. One the one hand, the benefits of such talker variability is supported in research and plays an important role in HVPT (Lively et al., 1993; Thomson, 2018b), to help improve generalizability of improvement from training to new talkers. However, many students strongly rely on their instructors as models of language and may receive an imbalance of input types throughout their learning (e.g., more of one type than another). Though this imbalance may seem inconsequential, research confirms that there is indeed a difference in listener perception of male and female talkers (Levitt & Lucas, 2018; Pépiot, 2013; Rodero, et al., 2013; Yiu et al., 2008). If there is a difference in perception, then it may be that sex of talker matters in the learning process. Therefore, it is important to investigate the impact of higher and low-pitched talkers in pronunciation acquisition to determine if more attention should be paid to talker quality in the learning process.

Research Questions

Utilizing the web application English Accent Coach (Thomson, 2012; Thomson, 2018a), this study strives to answer the following research questions:
1. Do participants acquire more accurate L2 perception:
   a. when utilizing an IPA or a Color Vowel Chart in HVPT?
   b. when exposed to male or female talkers?
2. Do participants acquire more accurate L2 production:
   a. when utilizing an IPA or colours in HVPT?
   b. when exposed to male or female talkers?

METHOD

Participants

Eight adult female English as a Second Language (ESL) learners enrolled in ESL classes at Brock University were recruited for this study. Participants were between the ages of 19-23, with the exception of one participant at 43 years of age. The learners were of mixed proficiency levels (low intermediate to high intermediate) and came from Japan (5), Korea (2), and China (1). Five participants reported having previous experience using the IPA.

Tests & Procedures

Testing Groups

Participants were assigned numbers and divided into four groups using a matched group design (see Figure 1). To account for certain participants having more or less knowledge of IPA and varying in L2 proficiency, the researchers placed participants in groups to attempt to balance these differences, as outlined in Table 1 below. As a result, four groups were created; two received training with male talkers with one utilizing IPA and the other utilizing colours, and two groups received training with female talkers with one using IPA and the other using colours.

Figure 1
Diagram of matched group design.
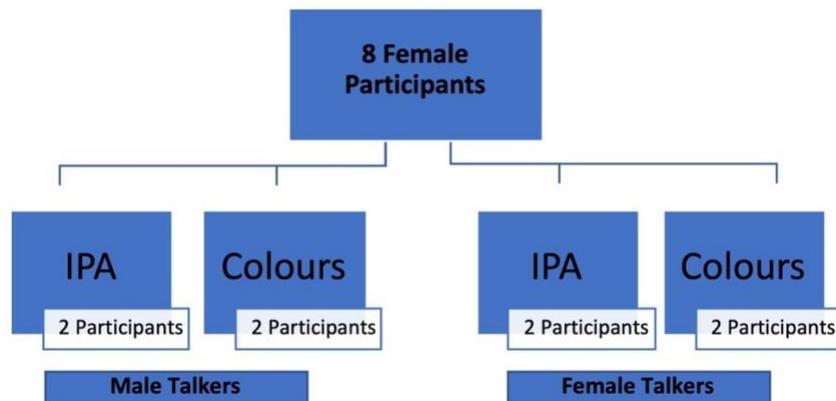
Table 1
Participant information and groupings based on IPA knowledge and proficiency level

Group	Participant	Age	L1	IPA Knowledge	Proficiency Level (1-5)
Group 1 (IPA; Male Talkers)	6	19	Japanese	No	Level 3
	3	19	Japanese	No	Level 2
Group 2 (Colour; Male Talkers)	8	22	Mandarin	Yes; Rating: 3	Level 4
	2	19	Japanese	No	Level 2
Group 3 (IPA; Female Talkers)	7	19	Japanese	Yes; Rating: N/A	Level 3
	10	23	Korean	Yes; Rating: 2	Level 4
Group 4 (Colour; Female Talkers)	4	43	Japanese	Yes; Rating: 2	Level 3
	9	21	Korean	Yes; Rating: N/A	Level 4

L2 Perception Familiarization Session (Pre-test) and Post-test

To investigate Research Question 1, a familiarization session was designed and used as a pre-test for perception, and then again as a post-test. Participants were first pre-taught certain elements of the web-based program and the logistics of each future session (e.g., how to login to the website, how to load a session, etc.). Participants then received initial training by practicing 100-word sounds represented in IPA and 100-word sounds represented in colours; both heard a mix of male and female talkers. The following vowel sounds were included in this training: [i, ɪ, e, ɛ, æ, a, ʌ, o, ʊ, u]. In this pre-test, the learners listened to recordings of English syllables beginning with /h/ followed by a vowel (i.e., /hV/) and identified the vowel sounds they heard by clicking on a symbol (IPA or colours) representing each vowel sound they perceived in each stimulus item (see Figure 2).

Figure 2
English Accent Coach forced choice ID paradigm, using either the International Phonetic Alphabet (left) or colours (right) as response symbols.
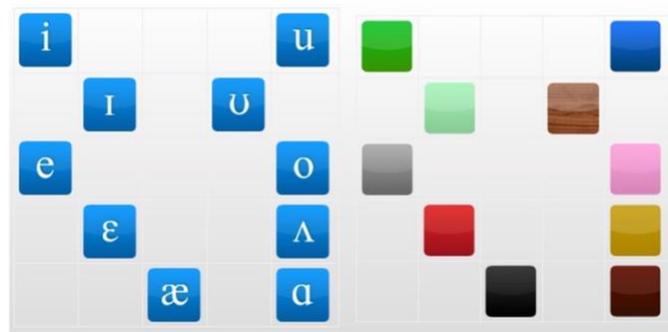
For example, if the word /ho/ was heard, participants would choose the symbol [o] displayed on the screen in the IPA condition, or they would choose the colour rose in the colour condition, because the sound /o/ is the vowel in the word r<u>o</u>se. Other colours included g<u>ree</u>n, m<u>i</u>nt, gr<u>e</u>y, r<u>e</u>d, bl<u>a</u>ck, <u>au</u>burn, m<u>u</u>stard, w<u>oo</u>d, and bl<u>ue</u>. Learners were given visual feedback after each response, indicating whether their choices were correct or incorrect. The perception post-test in Meeting 6 was identical to the pre-test familiarization session apart from the pre-teaching element, with participants receiving feedback on the accuracy of each response.

Production Pre- and Post-test

To investigate Research Question 2, a production pre- and post-test was designed; participants heard a recording of both a male and female talker using accents typical of the population in St. Catharines, Ontario. In the recording process, based on that of Thomson (2011), the speakers heard the target word presented in the carrier phrase, "The next word is__", and participants were prompted to reply by producing the target word in the carrier phrase "Now I say__". The production targets included /hV/, /bV/, and /hVd/ frames for each of the ten target vowels, isolating each vowel pronunciation in a relatively similar context (see Appendix A). Each participant undertook both the production pre- and post-tests individually (Meetings 1 and 7 respectively) and

took a total of approximately 6 minutes to complete them. After recording each speaker, Sound Studio 3 and Audacity were used to extract the target words and save them as individual sound files. All the participants' recorded productions were randomly evaluated by two listeners (the principal investigator and another phonetically trained listener) using Praat's Forced Choice Identification script.

Training Sessions

In Meetings 3-5, participants engaged in training sessions on 150-items, with 2 sessions per meeting. The training sessions ran similarly to the familiarization session and included the same phonetic environment (i.e., /hV/), with the only difference being the group conditions: instead of all participants hearing multiple male and female talkers and identifying sounds with both IPA and colour symbols, learners heard either multiple female or multiple male talkers, and responded using only IPA or only colour symbols, as required by their group assignment. The training sessions varied in duration for each participant, but none spent more than 40 minutes per meeting time. After the pre-test and familiarization session, there were a total of six training sessions with two sessions per meeting, as depicted in Figure 3, with an overall duration of three weeks to complete Meetings 1-7.

Figure 3
Order of events (Pre-tests, training sessions, and post-tests)
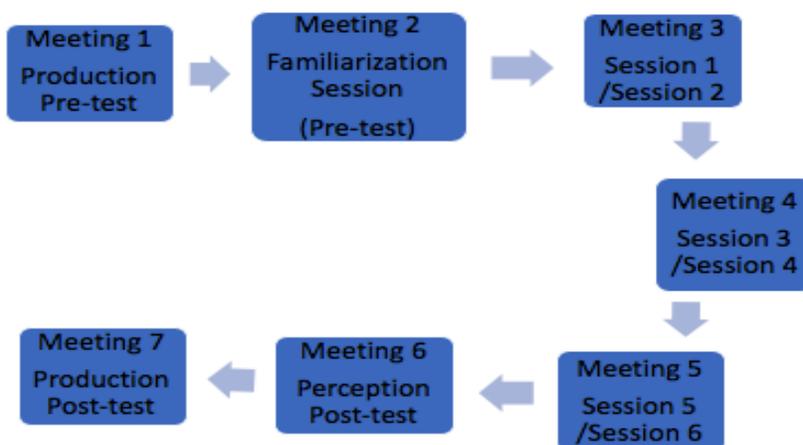
RESULTS

L2 Perception

The first part of Research Question 1 asked whether participants acquire more accurate L2 perception when utilizing IPA or colours in HVPT. Mean pre- and post-test scores for IPA and colours revealed that all participants improved, regardless of visual response buttons (see Table 2). In particular, the colour group made great gains from pre- to post-test, with both groups reaching similar accuracy scores at the post-test.

Table 2

Mean score of each Group in L2 perception pre- and post-tests for IPA and Colour groups

Pre-Test		Post-Test	
IPA Training	71.75%	**IPA Training**	79.75%
Colour Training	57.00%	**Colour Training**	78.25%

The second part of Research Question 1 investigated whether participants acquire more accurate L2 perception when exposed to male or female talkers. Pre- and post-test scores for male and female talker trained groups both showed that all participants improved, regardless of what talker type they were trained on (see Table 3). The data also reveals a potential training effect for the sex of talker: participants generally perceived sounds better in the post-test from talkers with the same sex as those that they were trained on. This became more evident when looking at the top five most difficult sounds to perceive for each participant. Even though both groups achieved overall perception gains, for 6/8 participants, their perception scores on the opposite sex talker from their training decreased by the post-test, with three participants in the male talker trained group getting worse at perceiving sounds from female talkers by the post-test (see Table 4), and vice versa for two of the participants in the female talker trained group (see Table 5).

Table 3

Mean score of each group in L2 perception pre- and post-tests for male and female talker groups

Pre-Test		Post-Test	
Male Talker Group	67.75%	**Male Talker Group**	80.00%
Female Talker Group	61.00%	**Female Talker Group**	78.00%

Table 4

Correct ID scores at pre- and post-test for male talker trainees X test talker sex for five most difficult vowels only

	Percent correct when test vowel produced by male talkers		Percent correct when test vowel produced by female talkers	
	Pre-Test	Post-Test	Pre-Test	Post-Test
Participant 6	60%	80%	40%	20%
Participant 8	20%	80%	60%	20%
Participant 3	50%	80%	25%	20%
Participant 2	20%	0%	60%	100%

Table 5

Correct ID scores at pre- and post-test for female talker trainees X test talker sex for five most difficult vowels only

	Percent correct when test vowel produced by male talkers		Percent correct when test vowel produced by female talkers	
	Pre-Test	Post-Test	Pre-Test	Post-Test
Participant 7	80%	20%	20%	40%
Participant 4	20%	40%	60%	60%
Participant 9	40%	20%	60%	80%
Participant 10	20%	40%	80%	40%

L2 Production

The first part of Research Question 2 investigated whether participants acquire more accurate L2 production when using IPA or colours in HVPT. Results indicated that both groups made gains in L2 production regardless of whether they were trained using IPA or colours (see Table 6).

Table 6
L2 production pre- and post-test scores when hearing both a male and female talker – IPA vs. colour trained group

IPA Training				Colour Training			
Participant #	Pre-Test	Post-Test	Average Gains	Participant #	Pre-Test	Post-Test	Average Gains
Participant 6	67%	75%	8%	Participant 8	65%	83%	18%
Participant 3	53%	63%	10%	Participant 2	58%	68%	10%
Participant 7	48%	57%	9%	Participant 4	65%	70%	5%
Participant 10	68%	73%	5%	Participant 9	73%	77%	4%
Average Gains	8.00%			Average Gains	9.25%		

The second part of Research Question 2, whether participants acquire more accurate L2 production when exposed to male or female talkers, seems to follow the same trend: participants acquired more accurate L2 production, regardless of which talker participants they were trained on (see Table 7). However, when looking at the gains from each group, there may be an opposite sex effect for those in the male talker training groups, with participants making more gains than those in the female talker-trained group. Given the small sample size, this is admittedly speculative.

Table 7
L2 production pre- and post-test scores when hearing both a male and female talker – male vs. female trained groups

Male Talker				Female Talker			
Participant #	Pre-Test	Post-Test	Average Gains	Participant #	Pre-Test	Post-Test	Average Gains
Participant 6	67%	75%	8%	Participant 7	48%	57%	9%
Participant 3	53%	63%	10%	Participant 10	68%	73%	5%
Participant 8	65%	83%	18%	Participant 4	65%	70%	5%
Participant 2	58%	68%	10%	Participant 9	73%	77%	4%
Average Gains	11.50%			Average Gains	5.75%		

DISCUSSION

In three weeks, participants not only showed gains in L2 sound perception with HVPT, but also in L2 production despite not having explicit practice. The L2 perception scores for IPA vs. colours improved for all participants, regardless of their training group. These results, in combination with past research (Thomson, 2011) suggest that the type of symbol might not matter in HVPT: learners will acquire more accurate L2 perception regardless, as long as they are mapping sounds to a consistent visual representation. Alhough it appears that the colour group experienced more gains

in L2 vowel perception than the IPA group, further analysis of each group's average score in the pre-test reveals that the colour group had a much lower average score initially. Therefore, it is unclear if the greater gain is due to the colour condition in the training itself, or due to the differences in initial scores.

Upon analyzing male vs. female talker group data for L2 perception, it is possible that there is a training effect for the sex of talker: participants generally perceived sounds better in the post-test from talkers with the same sex as those that they were trained on. Therefore, though L2 perception will improve overall, the sex of the talker that learners are trained on may have a role in determining L2 perception improvement.

In this study, L2 production was not explicitly taught to participants but was nevertheless investigated to confirm if a connection exists between perception and production. In accordance with previous research (Derwing & Munro, 2015; Lambacher et al., 2005; Qian et al., 2018), participants acquired more accurate L2 production simply because they were able to better perceive different sounds by the post-test. In the IPA vs. colours group data, just as with L2 perception, results showed that it might not matter if learners are trained on IPA or colours: they will acquire more accurate L2 production regardless of the response labels used.

For L2 production with different stimulus talkers, participants began with similar scores for the production pre-test, and yet, by the post-test, those trained on male talkers achieved double the gains than that of those trained on female talkers. This finding supports the conclusion from literature that higher-pitched talkers may reduce the intelligibility of vowels (Diehl, Lindblom, Hoemeke, & Fahey, 1996) and, given that all participants were female, this finding aligns with evidence that shows there may be an increase in attention and sensitivity to voices of the opposite sex (Junger et al., 2013), and that females in particular may have an increased sensitivity to low frequencies (Hunter et al., 2005). If there is indeed an increased sensitivity to voices of the opposite sex, then perhaps by being trained on male talkers, female participants gained an increased sensitivity advantage when relating the vowels they perceived to the vowels they intended to produce, leading to greater gains than the female talker trained group in the same amount of time. Since all groups made gains in perception and production, it is possible that the female talker trained group simply needed more time to reach similar production accuracy. However, given the very small sample size, these findings are speculative.

The current study is not exempt from limitations; first, the number of participants was not large enough to allow for statistical tests of significance. Second, all participants were female, and thus, the conclusions made cannot apply to male learners. Finally, participants did not have homogeneous backgrounds; L1, age, and proficiency level varied. Given the number of limitations, the question remains: How does this study inform both research and practice? Although multiple studies confirm the effectiveness of HVPT, the impact of talker quality on pronunciation learning is under-researched; given the results of this exploratory study, talker quality merits further investigation. For pronunciation teaching, this study introduces a new challenge – if talker quality is indeed a factor in learning, how do teachers adjust their instruction to include different talkers (male and female)? Though educators cannot control the type of talker students are exposed to on a day-to-day basis, they can incorporate individualized learning plans for pronunciation in the classroom utilizing HVPT; this would avoid a perception training effect with one type of voice

and provide students with more focused attention to phonetic differences. By utilizing digital resources such as English Accent Coach, teachers can further facilitate pronunciation acquisition for L2 learners and add a valuable, technological tool to the language classroom.

ACKNOWLEDGEMENTS

ABOUT THE AUTHOR

Kelsey Labbé works as the Academic Support Specialist with Brock International at Brock University. She enjoys all aspects of language learning, but in particular, takes an interest in pronunciation learning, sociolinguistics, and the accessibility of research to teachers in practice. Apart from her time in academia, she also enjoys pursuing her passion for music as a musician in the Canadian Armed Forces with the Naval Reserve Band at HMCS STAR in Hamilton.

REFERENCES

Breitkreutz, J.A., Derwing, T.M., & Rossiter, M.J. (2001). Pronunciation teaching practices in Canada. *TESL Canada Journal, 19*(1), 51–61. https://doi.org/10.18806/tesl.v19i1.919

Brinton, D. M. (2017). Innovations in pronunciation teaching. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (1st ed., pp. 449-461). New York: Routledge.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research.* Amsterdam: John Benjamins.

Diehl, R.L., Lindblom, B., Hoemeke, K.A., & Fahey, R.P. (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics, 24*, 187-208. https://doi.org/10.1006/jpho.1996.0011

Hunter, M.D., Phang, S.Y., Lee, K.H., & Woodruff, P.W.R. (2005). Gender-specific sensitivity to low frequencies in male speech. *Neuroscience Letters, 375*(3), 148-150. http://doi.org/10.1016/j.neulet.2004.11.003

Iverson, P., Pinet, M., & Evans, B.G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics, 33*(1), 145-160. http://doi.org/10.1017/S0142716411000300

Jones, T. (2017). Pronunciation with other areas of language. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (1st ed., pp. 370-384). New York: Routledge.

Junger, J., Pauly, K., Bröhr, S., Birkholz, P., Neuschaefer-Rube, C., Kohler, C., Schneider, F., … Habel, U. (2013). Sex matters: Neural correlates of voice gender perception. *NeuroImage, 79*, 275-287. http://dx.doi.org/10.1016/j.neuroimage.2013.04.105

Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels

by native speakers of Japanese. *Applied Psycholinguistics, 26*(2), 227-247. https://doi.org/10.1017/S0142716405050150

Levitt, A., & Lucas, M. (2018). Effects of four voice qualities and formant dispersion on perception of a female voice. *Psychology of Language and Communication, 22*(1), 394-416. http://doi.org/10.2478/plc-2018-0018

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3), 1242-1255. https://asa.scitation.org/doi/10.1121/1.408177

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research, 50*(6), 1496- 1509. https://doi.org/10.1044/1092-4388(2007/103)

Pépiot, E. (2013). Processing female and male voices: A word spotting experiment. *Perceptual & Motor Skills,* 117(3), 903-912. http://doi.org/10.2466/24.27.PMS.117x31z7

Qian, M., Chukharev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning & Technology, 22*(1), 69-96. http://doi.org/10125/44582

Rodero, E., Larrea, O., & Vázquez, M. (2013). Male and female voices in commercials: Analysis of effectiveness, adequacy for the product, attention and recall. *Sex Roles, 68*(5-6), 349-362. http://doi.org/10.1007/s11199-012-0247-y

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal, 28*(3), 744–765. http://doi.org/10.11139/cj.28.3.744-765

Thomson, R. I. (2012). English Accent Coach: Not quite a fairy godmother for pronunciation instruction, but a step in the right direction. *SHARE*, *4*, 13-17. https://www.tesl.ca/images/docs/share/share_4_2013.pdf

Thomson, R. I. (2018a). English Accent Coach [Computer program]. Version 2.3. www.englishaccentcoach.com

Thomson, R.I. (2018b). High variability [pronunciation] training: A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation, 4*(2), 208-231. https://doi.org/10.1075/jslp.17038.tho

Thomson, R.I., & Derwing, T.M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*(3), 326–344. https://doi-org.proxy.library.brocku.ca/10.1093/applin/amu076

Yiu, E.M.L., Murdoch, B., Hird, K., Lau, P., & Ho, E.M. (2008). Cultural and language differences in voice quality perception: A preliminary investigation using synthesized signals. *Folia Phoniatrica et Logopaedica, 60*(3), 107-119. http://doi.org/10.1159/000119746

Appendix A

List of word sounds heard in the production pre- and post-test

1. hi	11. bi	21. heed
2. hɪ	12. bɪ	22. hid
3. he	13. be	23. hayed
4. hɛ	14. bɛ	24. head
5. hæ	15. bæ	25. had
6. ha	16. ba	26. hawed
7. hʌ	17. bʌ	27. hud
8. ho	18. bo	28. howed
9. hʊ	19. bʊ	29. hood
10. hu	20. bu	30. who'd