Summer 8-15-2020

Adding Production to High Variability Phonetic Training

Caleb Crosby

ADDING PRODUCTION TO HIGH VARIABILITY PHONETIC TRAINING

by
Caleb Crosby

A thesis submitted to the faculty of The University of Mississippi in partial fulfillment of the requirements of the Sally McDonnell Barksdale Honors College.

Oxford
August 2020

Approved by

____
Advisor: Dr. Vance Schaefer

____
Reader: Dr. Ala I. Simonchyk

____
Reader: Dr. Tamara Warhol

ACKNOWLEDGEMENTS

This research project would not have even been started without the influence of the community of student-teachers, volunteers and learners involved in the Community ESL classes at the University of Mississippi's Intensive English Program, who showed me how engaging teaching ESL can be, and convinced me that teaching is *not* easy and methods deserve to be researched, or without the professors who took genuine interest in my well-being, encouraged, and guided me as I stumbled through my first semesters at the University of Mississippi. I would also like to thank the research participants who graciously took time out of their own busy schedules to take part in the study. This project would have not been possible without the guidance of my thesis advisor, Dr. Vance Schaefer, who first sparked my interest in teaching pronunciation. His rigorous questioning during the conception of the project and endless patience during setbacks were beyond value.

ABSTRACT

The effectiveness of adding a production component to a High Variability Phonetic Training (HVPT) regimen to improve native Japanese speaker's pronunciation of English [b], [v], [f], and [h] was investigated. L1 Japanese-speaking English learners were recruited as participants, and a pretest-posttest procedure was used to evaluate improvement at production of the target consonants. For the pretest and posttest, recordings were taken of participants pronouncing twelve tokens, and the recordings were rated for intelligibility by a phonetically trained native English-speaking rater. Participants were divided into two groups. Group A received only HVPT training, and group B received a regimen of half HVPT training and half production practice. Performance during the HVPT portions of the training was tracked, and the pretest and posttest were compared to determine if improvement at production of target consonants occurred. Although findings were largely inconclusive, clear patterns emerged that may offer insight into how native Japanese speakers perceive particular sound contrasts in English.

TABLE OF CONTENTS

I. Literature Review

Most adult second language learners do not develop native-like pronunciation of vowels and consonants in their second language (Munro & Derwing, 2008). Yet for many second language learners, improving pronunciation is still an important goal. Derwing and Rossiter (2002) interviewed 100 immigrants attending full time classes in an ESL program in Canada, and found that more than half (55%) of the students felt that pronunciation problems contributed to communication difficulties they had when speaking English. Many (42%) said that pronunciation was the primary cause of those difficulties. Derwing (2003) conducted another study that investigated adult immigrants' perceptions of their own pronunciation problems and the consequences of speaking with a foreign accent. More than half of the participants felt that pronunciation played a role in their communication problems and that people would respect them more if they had better English pronunciation, yet when asked what aspect of their pronunciation needed to be improved, many were unable to answer.

Students' impressions of the importance of pronunciation and communication ability are not unfounded. Being able to communicate effectively and relate with people in English is very important for the social well-being of adult immigrants to English-speaking countries (Derwing, Thomson, & Munro, 2006). Listeners may react negatively to hearing a speaker's accent and react based on their prejudices (Lippi-Green, 2012), even going so far as to harass, refuse employment, or deny housing based solely on their speech (Munro, 2003).

Listener prejudices aside, intelligibility defined as “the extent to which a speaker’s message is actually understood by a listener” (Derwing & Munro, 2015) is crucial. Whether in a life and death situation, such as pilots and co-pilots receiving flight and landing instructions from ground control, diplomats negotiating an international treaty, or doctors and nurses communicating a patient’s condition, or less serious situations like talking to family and friends over dinner, people want to understand and be understood. However, for second language learners, achieving comprehensible pronunciation can be time consuming and frustrating.

If a person wants to be able to communicate in a foreign language, they must learn to recognize and produce the sounds of that new language, i.e. they must perceive segments (vowels and consonants), suprasegmentals such as English stress (how loud and long each syllable is), and phonotactics (what sounds can appear next to each other and what order they can appear in). However, the task is a little bit more complicated than simply paying attention to words, trying to remember the sounds, and practicing moving your mouth in the right way. Goto (1971) tape-recorded eight Americans and eleven Japanese participants, then tested those same eight Americans and eleven Japanese to see if they could distinguish which words contained /l/ and which ones contained /r/. Goto found that native speakers of Japanese who learn English as an adult still have difficulty perceiving the acoustic differences between English /r/ and /l/ whether they were listening to the voices of the Americans or their own voices. So even if a learner tries to listen carefully and learn how to pronounce their new language, they may not be able to accurately hear the differences between some sounds.

Second language pronunciation is thus difficult because learners can not rely on learning to listen to and pronounce a second language the same way they learned their first. In fact, the ability to distinguish non-native contrasts is greatly diminished by the time we are adults. In one study, Werker, Gilbert, Humphrey, and Tees (1981) compared L1 English infants, L1 English adults, and L1 Hindi adults on their ability to discriminate between two pairs of Hindi speech contrasts not found in English. The study indicated that infants are able to discriminate speech sounds according to phonetic category without prior specific language experience, but that adults have much difficulty discriminating sound contrasts not found in their native language. Werker and Tees (1984) compared English-speaking adults, infants learning English, and Thompson (a Native American language) speaking adults, and demonstrated that infants around ages 6 to 8 months could distinguish the Thompson contrasts as could the Thompson-speaking adults, but that the L1 English-speaking adults could not. They had lost sensitivity to the non-native sound contrasts.

Humans use language as a tool for communicating and transmitting information, and normally pay more attention to the meaning of sounds than minor variations in the sounds themselves, which have been said to be perceived by native speakers as belonging to mental categories (Liberman et al, 1957). According to the Perceptual Assimilation Model (Best, 1995), adults who are listening to language sounds hear those sounds according to how acoustically similar or dissimilar they are to the phonemes found in their native language. That is, these phonemes form abstract categories. For example, the /t/ in ‘top’ and ‘stop’ are physically different: the first is aspirated (air comes rushing out of the mouth when it is pronounced) and the second is unaspirated (very little air comes out when it is pronounced). They are both allophones (different members of a set or category that are recognized by native speakers as being the same thing) of the phoneme /t/, and recognized by L1 speakers of English as /t/, but are physically different.

Thus, when L1 speakers of English learn a language where aspiration is phonemic (where aspiration can be used to distinguish meaning), they may have difficulty establishing a new category for unaspirated stops. Best’s (1995) PAM model explains this process: Two categories in the new language could be overlapped by one category in the learner’s native language, leading to difficulty distinguishing the sounds and remembering which one to pronounce in a given word. A well-known documented example of this is the mapping of English /l/ and /r/ onto a flap /ɾ/ by L1 Japanese speakers when learning English (Goto, 1971). Additionally, one category in the second language could map onto two categories in the learner’s native language, leading to the learner’s confusion as to why sounds are different but treated the same. A new sound also may not have any corresponding phoneme in the learner’s native language, for example, Zulu clicks (Best, McRoberts, & Sithole, 1988).

While PAM is used to account for how native language shapes perception of consonants and vowels by functional, monolingual adults, the Speech Learning Model

(SLM) (Flege, 1995) and the Perceptual Assimilation Model of Second Language Speech Learning (PAM-L2) (Best and Tyler, 2007) account for how L2 learners form new categories for sounds in their L2. Both models assume that a learner has one perceptual system that is used for all of the learner's languages. If an existing category overlaps two phonemic categories in the learner's L2, perceptual learning will have to take place for the learner to distinguish the two new categories. SLM also assumes that speech production will eventually come to match speech perception. That is, learners will eventually produce output that matches what they perceive; thus, accurate perception is critical to developing accurate pronunciation.

Despite the fact that much of the considerable body of research on L2 speech is theoretical in nature and not directly applicable to language teaching (Derwing & Munro, 2015), pronunciation instruction has been found to help learners improve their pronunciation (Lee, Jang, & Plonsky, 2014). A major push for native-like pronunciation came with the Audiolingual Method of teaching, which appeared in the mid-20th century and was especially popular in North America (Derwing & Munro, 2015).The Audiolingual Method required learners to listen to and imitate native speaker models as closely as possible to develop target-like pronunciation. In the later part of the 20th century, Communicative Language Teaching moved away from the Audiolingual Method, and de-emphasized pronunciation due to it being considered unteachable and because it was believed that learners would acquire whatever skills they needed through simple exposure and practice (Derwing & Munro, 2015).

However, listeners who are presented with speech stimuli from phonetic categories that are not present in their native language will typically perform much worse at identifying those sounds than a native speaker of the language from which the stimulus phonemes were selected (Werker & Logan, 1985). Categorical perception and the accompanying loss of sensitivity to L2 phonemic contrasts in adults, as discussed earlier, implies that most students will probably need some modification of their perception, or perceptual learning, in order to see maximum gains in pronunciation accuracy.

1.1 Historical Context for High Variability Phonetic Training

In language classes, identification, or perception training with feedback has been demonstrated to improve learner performance in the identification of sounds, leading to some pronunciation gains compared to untrained learner control groups (Lambacher et al., 2005). However, in language classrooms, input is typically provided by the teacher, or a single voice on an audio tape or CD (Barriusso & Hayes-Harb, 2018). So even if learners demonstrate improvement inside the classroom, the instructor cannot be certain that learning will generalize to real life situations outside the classroom. Exposure to only a few voices with clear pronunciation may not be sufficient to prepare learners for the range of phonetic variation used by other speakers they encounter, or the variety of listening conditions, such as background noise in restaurants, street traffic, other conversations nearby, that they may find themselves in.

Because variation within categories is natural, Thomas & Derwing (2016) have suggested that teaching the pronunciation of L2 sounds should incorporate and emphasize variation rather than focusing on prototypes and citation forms, which are difficult to imitate perfectly, or the pronunciation of sounds in isolation. One technique that allows for a controlled approach to input variability is High Variability Phonetic Training (HVPT).

HVPT is a method of perceptual training that uses audio recordings from a variety of speakers to deliver stimuli to a learner. The learner responds to the stimuli by identifying which sound they hear in a designated part of the stimuli, and receive immediate feedback on the correctness of the response. Since HVPT provides learners with feedback on the accuracy of their perceptions, it can help to direct their attention to the properties of segmental stimuli that are important for L2 category formation (Thomson & Derwing, 2016). That is, learners are trained to pay attention to the acoustic properties that are important for distinguishing phonemes (i.e. use formants to identify particular vowel sounds), and to ignore the different properties that vary by speaker.

Logan, Lively, and Pisoni (1991) found that laboratory training procedures can be used to modify Japanese listeners' perception of /r/ and /l/ in isolated English words. Six native speakers of Japanese who had been living in the United States for periods ranging from 6 months to 3 years were trained and evaluated using a pretest-posttest procedure. Participants were exposed to /r/-/l/ minimal pairs with those sounds in word initial and word final positions, as well as in single consonant, consonant cluster, and intervocalic configurations. Six talkers, four male and two female, recorded the stimuli. During the pretest, posttest, and training, participants were required to listen to the word and mark which word they heard in an answer book provided to them. A significant improvement was seen between the pretest and the posttest. Interestingly, the participants' performance in identifying /r/ and /l/ depended on whether the participant had heard that speaker before or not. This suggests that if researchers want to ensure that results are generalizable, we should not use tests where participants listen to speakers that they have been exposed to already.

Lively, Logan, and Pisoni (1993) replicated their study and added a task where participants were given new words produced by a familiar and an unfamiliar speaker. Participants were able to accurately identify sounds from both speakers, showing that their learning had been generalized. However, Thomson and Derwing (2016) found that when the stimuli contain mainly real words, learners do not improve their performance in spontaneous production tasks.

Thomson and Derwing (2016) investigated whether perceptual training using nonsense words or training that predominantly focused on real words resulted in better pronunciation of real words. Their participants were divided into three groups, including a control group. The first experimental group was a Phonetic Group, who were given training for English vowels inside isolated open syllables (e.g., /bi/, /pi/, bɪ/ and /pɪ/), which are mostly not real words, forcing participants to pay attention to phonetic values. The second experimental group was a Real Word Group, who were almost entirely given the target words. In the context of real words, participants are often able to recognize the word, and may know which vowel is “supposed” to be inside the word instead of focusing on its phonetic qualities.

Participants in both experimental groups completed 40 training sessions at their leisure over the course of one month, but were told that they could complete a maximum of two sessions per day. Between pretest and posttests, 89% of the Phonetic Group improved over time, however only 50% of the Real Word Group and 60% of the Control Group demonstrated improvement in their mean scores. Due to these small gains, Thomson and Derwing (2016) suggested that perceptual training on its own is insufficient to promote maximal improvement and that learners probably need to practice producing the sounds they are learning to more accurately perceive. They also suggested that some English sounds appear to be easier for learners because either there are direct parallels in their L1, or because they may simply be easy to perceive and produce. This raises the question, what will be the effect on pronunciation if a production component is added to perception training with whole words?

According to an ESL teacher who is experienced at teaching English to native Japanese speakers, [f]/[h], [b]/[v], and [r]/[l] are categories that are very challenging for native Japanese speakers (Nakayama, 2019). Among these, the [r]/[l] contrast has been heavily studied, however [f]/[h] and [b]/[v] have gotten much less attention. For these reasons, this study focuses on the pronunciation of English [f]/[h] and [b]/[v] contrasts among native speakers of Japanese.

1.2 Summary of HVPT benefits for learners

1. Learners are exposed to a variety of speakers.

2. Learners get immediate feedback on their perceptions which facilitates L2 category formation.

3. Improved L2 category formation leads to improved perception of L2 segments.

4. Improved perception leads to pronunciation gains.

1.3 Research Questions

In response to this, the current study asks the following questions:

Research Question 1:

Does adding a production component to perception training with whole words improve pronunciation more than perception training alone?

Research Question 2: Does production practice improve scores on intelligibility in English among native speakers of Japanese, as rated by native or proficient English speakers, when the participants are producing English words that contain the sounds [f], [h], [b], or [v]?

II. Methods

The following methods were used to conduct this study:

2.1 Participants

Seven participants completed at least one session of the treatment. They were recruited by speaking with Japanese exchange students. The participants were told that the researcher is conducting a small study as part of the graduation requirements for the honors college, and that the project is investigating methods to improve English pronunciation among native speakers of Japanese. Five participants were graduate students, and two were students in an Intensive English Program. Although all participants completed a pretest, the number of sessions that each participant was able to complete varied. One participant was able to complete only one session before returning to Japan. Three participants completed two sessions of treatment, and three participants were able to complete all three sessions of HVPT and the posttest. Various environmental factors, including the influence of the novel coronavirus (COVID-19), made it difficult to ensure that all participants completed all parts of the study under strictly the same conditions. There were also time intervals of varying size between each session due to the individual nature of each participant's schedule. The entire recruitment process was conducted in accordance to the Institutional Review Board (IRB) approval (See Appendix B for the consent form).

The rater was a native speaker of Southern American English and was phonetically trained in a graduate seminar for second language phonology where research concerning L2 pronunciation was covered as part of the course.

2.2 Tasks

Two sets of tasks were conducted. One set consisted of only listening/discrimination training with feedback. The other set consisted of listening/discrimination training with feedback immediately followed by production practice with feedback. As such, participants were first divided into two groups: group A and group B. Both groups were given a pretest consisting of recording the pronunciation of twelve words. Group A received only listening/discrimination training with feedback, and group B received listening/discrimination training with feedback immediately followed by production practice with feedback. Both groups received training for both [b]/[v] and [f]/[h] sound contrasts.

Group A participated in training sessions of 200 words per consonant pair of HVPT training using the English Accent Coach (EAC) website, while the HVPT portion of Group B's sessions consisted of only 100 words. Both groups were guided through using the EAC website for the listening/sound discrimination.

EAC is a free browser based HVPT training application that can be used to present a learner with auditory stimuli consisting of words or word fragments containing target vowels and consonants, recorded by thirty speakers of Canadian English, which is similar to General American (Thomson & Derwing, 2016). Learners interact with a simple interface, shown in Figure 1, that allows them to select a response by clicking the appropriate button with a computer mouse to indicate their response.

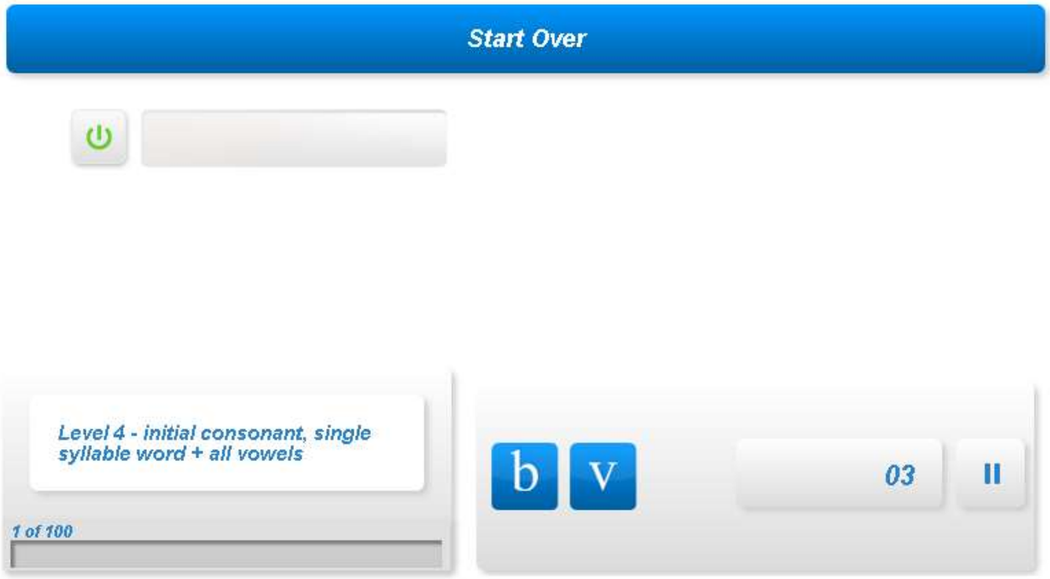
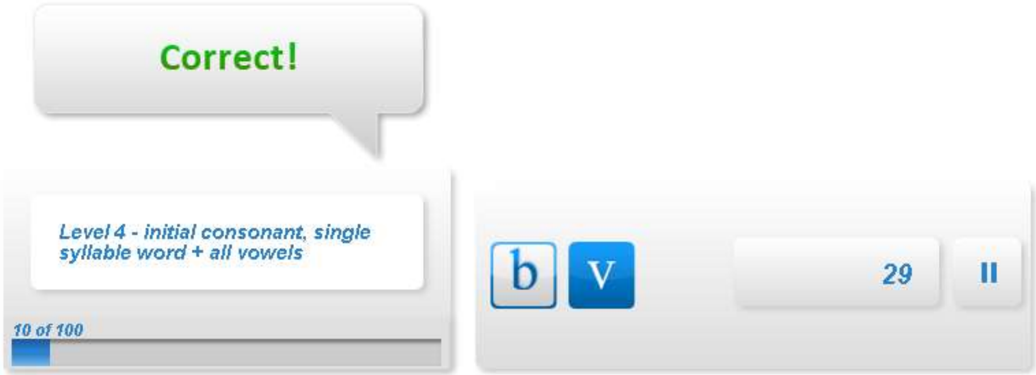

Start Over
Level 4 - initial consonant, single syllable word + all vowels
1 of 100
b
v
03

Figure 1: EAC interface

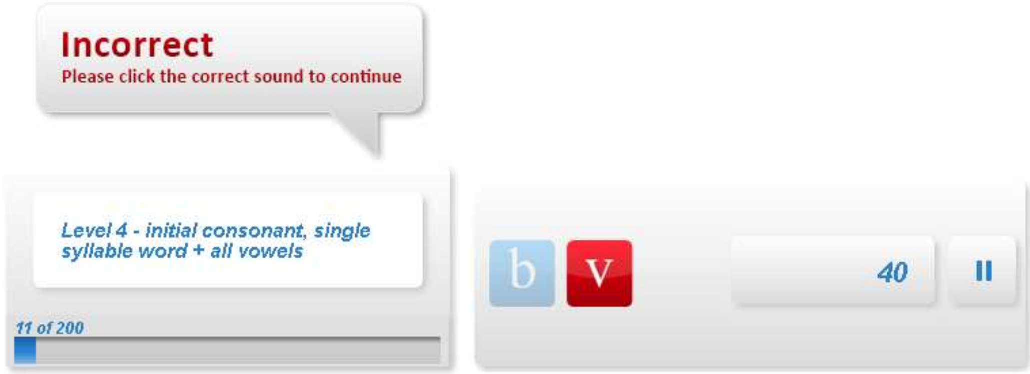
Participants use headphones connected to the computer to listen to an auditory stimulus which plays only once. Participants are then required to respond to each stimulus item by clicking on a button to identify the initial consonant. For the [b]/[v] section, a “b” button and a “v” button are available; for the [f]/[h] section, an “f” button and an “h” button are available. After making their choice, they receive auditory and visual feedback on the accuracy of their selection. If the selection is incorrect, the feedback includes a repeat playing of the stimulus word, and participants are prompted to select the correct answer. An example of the visual feedback that would be provided if a participant correctly selected “b” during the [b]/[v] portion of the training can be seen in Figure 2, and an example of the visual feedback for an incorrect answer of “b” can be seen in Figure 3.



Correct!

Level 4 - initial consonant, single syllable word + all vowels

Figure 2: Correct response



Incorrect
Please click the correct sound to continue

Level 4 - initial consonant, single syllable word + all vowels

Figure 3: Incorrect Response

Immediately following the HVPT portion, group B conducted a second portion that consisted of 100 words of production practice for both [b]/[v] and [f]/[h] contrasts, guided by a stimulus prompt. Participants sat at a table across from the researcher and read each word from the prompt one at a time. If the participant produced the word in a way that the researcher deemed to be non-target-like, they were stopped by the researcher asking for them to "please repeat the word." Participants were given no instruction on how to pronounce the word.

Participants who completed three training sessions participated in a posttest at the conclusion of the third training session (See Appendix D for the production stimuli used in the pretest and posttest).

2.3 Stimuli

The stimuli consisted of single syllable words that begin with the target sound, with no restrictions on what vowels can form the nucleus of the word. Some minimal pairs were included. Words were shuffled and produced at random by the EAC application, with some considerable repetition of words. Each training session was divided into two sets. Each word in the first set was a single syllable word that began with either the consonant [b] or [v], and each word in the second set was a single syllable word that began with either the consonant [f] or [h]. Table 1 shows a sample of the stimuli produced by EAC.

Table 1. Sample of Stimuli Produced by EAC Application

[b]/[v]	[f]/[h]
Vote	Feared
Boat	Hoed
Bob	Hung
Best	Heed
Bush	Hard
Vest	Whose
Booed	Feed
Bide	Had

The production stimuli were developed by taking a sample of stimuli produced by the EAC application to ensure that the production stimuli were similar in nature and the frequency of repetition of individual words to that of the HVPT portion of the training.

Table 2. Sample of Production Stimuli

[b]/[v]	[f]/[h]
Vast	Facts
Bid	Hope
Beth	Herd
Bath	Fast
Voice	Fault
Vote	Force
Bide	Find
Book	Hacked

2.3.1 Pretest and Posttest

Both the posttest and pretest prompts contained twelve words that were similar in structure (single syllable words where the target sound is the initial consonant) to both the HVPT and production practice stimuli. These items were taken from a sample of EAC stimuli just as the production stimuli were, but avoided minimal pairs. Pretest and posttest items were randomized and printed in a numbered list in Arial size 12 font, double spaced, to be read by the participants. Participants were given the pretest or posttest prompt and asked to read the list aloud with a short pause between each word. Participants were not allowed to practice the test items before the test. Twelve recordings per participant were produced by using an LG Phoenix 3 (LG-M150) cell phone and its default audio recorder application to digitally record audio at 128 kbps (similar to CD quality). Recordings were divided by stopping and restarting the recorder after each item on the test prompt.

Then, all the pretest and posttest recordings were randomized to be presented to the rater to be judged. The rater made use of a three-point system, where 3 points indicates that the participant's pronunciation was a good example of typical pronunciation for that phoneme, 2 points indicates that it was a poor example, and 1 point indicates that the pronounced sound was from a completely different phonemic category.

There was a total of 120 audio files included in the evaluation (seven pretests and three posttests with twelve recordings each), which each consisted of a single word. All rating was conducted in a single session by a single rater which may have resulted in some variation in scoring. While the use of a single rater is a shortcoming, the simple system which was used to rate them and the relatively small number of samples should have helped alleviate this a little. In addition, the pretest-posttest results were also considered with 3-point and 2-point categories collapsed. This shows approximately what the results would look like if the rater engaged in an identification task rather than a rating task, which reduces the subjectivity of the results by placing emphasis on whether or not the sample was intelligible.

2.4 Procedure

1. The researcher recruited potential participants and a rater. Participants were recruited from among native Japanese speakers who are students at the University of Mississippi. All potential participants were informed of risks associated with the study, their right to end participation at any time, and signed a consent form.
2. Participants took the pretest at the beginning of the first training session. Recordings were made of participants reading the pretest stimuli and those recordings were stored on an encrypted USB flash drive.
3. Treatment. Participants were divided into two groups. Experiment group A received treatment A, consisting of 200 words of HVPT for both [b]/[v] and [f]/[h] contrasts each session. Group B received treatment B, consisting of 100 words of HVPT and 100 words of production practice for both [b]/[v] and [f]/[h] contrasts.

4. Posttest. The posttest was administered at the end of the third session and was equivalent but not identical to the pretest.
5. The rater listened to pretest and posttest recordings and judged the intelligibility of the speaker in each one. All recordings and data were stored on an encrypted USB flash drive.
6. The rater sent the rating information to the researcher, and the researcher analyzed the results.

III. Results

Table 3 shows the average percentage of correct answers by phoneme across all HVPT sessions for all participants, including those who did not complete the posttest. We can see that participants had the worst performance at identifying [v], and much better at for [h], [b], and [f].

Table 3. Average % correct by phoneme across all HVPT sessions

Phoneme	Average % correct
[b]	95.9%
[v]	79.5%
[f]	98.6%
[h]	94.4%

Only two participants from group A and one participant from group B were able to receive the posttest, so only their data was used for the pretest-posttest comparison and evaluation portion of the study. Table 4 shows pretest and posttest average scores by phoneme for each participant who took the posttest, and each participant's improvement at that phoneme. Numbers in parentheses show what the averages would be if 3-point rated words were collapsed into the 2-point category.

Table 4. Pretest vs Posttest Average Scores by Phoneme

Participant Number	Group	Phoneme	Avg. pretest rating	Avg. posttest rating	Improvement
2	A	[b]	3 (2)	3 (2)	0 (0)
	Graduate	[v]	2.33 (1.66)	2.66 (2)	0.33 (0.34)
		[f]	2.33 (2)	3 (2)	0.67 (0)
		[h]	3 (2)	3 (2)	0 (0)
3	A	[b]	2.66 (2)	3 (2)	0.34 (0)
	IEP	[v]	1.66 (1.33)	2 (1.66)	0.34 (0.33)
		[f]	3 (2)	2.66 (2)	- 0.34 (0)
		[h]	3 (2)	3 (2)	0 (0)
6	B	[b]	3 (2)	3 (2)	0 (0)
	Graduate	[v]	2.66 (2)	2.66 (2)	0 (0)
		[f]	2 (1.66)	3 (2)	1 (0.34)
		[h]	3 (2)	3 (2)	0 (0)

Participants in both groups showed overall gains in pronunciation, with the exception of the pronunciation of the [f] phoneme by participant 3 in group A. The largest improvement was made in the pronunciation of the [f] phoneme by participant 6 in group B.

Table 5 shows average scores and score increase by phoneme across all participants on the pretest and posttest. We see that generally [v] and [f] received the lowest scores on the pretest, with much higher scores for [b], and perfect scores for [h]. Scores for [f] improved the most, with much less improvement for [b] and [v]. [h] did not show any improvement due to participants receiving perfect scores on both the pretest and posttest for that phoneme (See Appendix D for a complete list of test scores by individual item).

Table 5. Average Score Increase by Phoneme

		posttest	
phoneme	pretest score	score	increase
[b]	2.88 (2)	3 (2)	0.12 (0)
[v]	2.22 (1.66)	2.44 (1.88)	0.22 (0.22)
[f]	2.44 (1.88)	2.89 (2)	0.45 (0.12)
[h]	3 (2)	3 (2)	0 (0)

IV. Discussion

The goal of the study was to determine whether adding a production component to an HVPT regimen improved pronunciation more than an HVPT regimen alone. In a comparison of the increase in scores between the pretest and posttest by group A and B, both groups made improvement. The largest gain for any single phoneme was in group B (the production group), whose rating on [f] averaged an entire point higher on the posttest as compared to the pretest. However, one participant in group A actually performed worse at the [f] phoneme on the posttest than on the pretest. This suggests that production practice may be necessary to reliably maximize improvement.

In response to Research Question 1: Does adding a production component to perception training with whole words improve pronunciation more than perception training alone? Overall, all three participants who took the posttest showed improvement, and neither group showed drastically more improvement than the other, which suggests that the answer may very well be "no." However, the fact that one member of group A performed worse after training on the [f] phoneme suggests that pronunciation training may be more reliable with the addition of a production component, and that production may be necessary to see maximum improvement.

In response to Research Question 2: Does production practice improve scores on intelligibility among native speakers of Japanese, as rated by native or proficient English speakers, when the participants are producing words that contain the sounds [f], [h], [b], or [v]? Yes, it does. The data collected during this study supports the idea that perceptual training with or without production practice has an overall positive influence on pronunciation as measured by a test which uses a reading prompt. However, only group A, which received exclusively perceptual training with no production practice, showed an actual decrease in scores on the posttest. As such, pronunciation training programs should require a production component for greater efficacy in boosting learner's intelligibility.

Differences in rates of improvement between the phonemes can be understood in the context of PAM-L2 (Best & Tyler, 2007) which explains that learners will have a more difficult time perceiving and distinguishing phonemes that do not contrast in their native language and Speech Learning Model (Flege, 1995) which states that speech production will change over time to match perception. If we consider participant's HVPT scores alongside the pretest and posttest test scores, it can be seen that the phonemes that are more easily perceived by learners were the ones most likely to improve on the posttest. Participants performed highest on average at recognizing [f] (98.6%), which was also the phoneme that saw the highest average improvement (+0.45 points). Similarly, participants performed lowest at recognizing [v] sounds as compared to the other phonemes, and [v] was more resistant to improvement on average (+0.22).

There were also the following limitations to the study:

1. There were too few participants for any results to be statistically significant.
2. The coronavirus outbreak, among other environmental factors such as the lack of a dedicated facility for collecting data, impeded the collection of posttest data, and

the conditions for the pretest, posttest, and individual training sessions were not identical.

3. The pretest and posttest used a reading prompt, which is easy for the researcher to administer but may not accurately reflect the learner’s speech in real world communication (Thomson & Derwing, 2016).
4. The schedule for the practices was constrained by the participants' own schedules, so times between practices varied widely among participants. This, of course, allows for numerous confounding variables that cannot be accounted for.

Future Studies

The results of this study are good news for foreign language teachers as it is possible for students to make progress with practice, even if they are already at an advanced level. HVPT shows a lot of promise as a method for improving learner’s pronunciation because it can be seen that better perception leads to gains in pronunciation.

However, the future studies would benefit from the inclusion of an increased sample size, multiple raters, other methods of measuring improvement that do not use a reading prompt and can more accurately estimate production in real world conditions, a more controlled environment for testing and training, and more control over the training schedule.

Also, the utility of HVPT needs to be verified in more realistic classroom settings. One of the requirements for HVPT is that the participant is engaged and listening, and not simply guessing at the answer or clicking through to finish the task as quickly as possible. There were a few occasions during the study that participants expressed frustration or boredom at the training exercises. Thus, a plain HVPT regimen by itself, or with production practice where students merely read out loud, may not be appropriate for an educational setting where students are not personally invested in the material (i.e. mandatory education). However, the principles behind HVPT, that is variation in input and frequent immediate feedback for learners, should be kept in mind during the search for tools and methods to assist students.

V. References

Barriuso, T. A., & Hayes-Harb, R. (2018). High Variability Phonetic Training as a Bridge from Research to Practice. *The CATESOL Journal*, *30*(1), 177–194.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Baltimore, MD: York Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 345–360.

Best, C.T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 13–34). Amsterdam: John Benjamins.

Derwing, T. M. (2003). What do ESL students say about their accents? *The Canadian Modern Language Review, 59*(4), 547-566.

Derwing, T. M., & Munro, M. J. (2015) *Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research*. Amsterdan: John Benjamins Publishing Company.

Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System, 30*(2), 155-166.

Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System, 34*(2), 183-193.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233–277). Timonium, MD: York Press.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r." *Neuropsychologia, 9*(3), 317–323.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & 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.

Lee, J., Jang, J., & Plonsky, L. (2014). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345-366.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358.

Lippi-Green, R. (2012). *English with an accent: Language, ideology, and discrimination in the United States* (2nd ed.). London, UK: Routledge.

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*, 1242.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874-886.

Munro, M. J., & Derwing, T.M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, *58*(3), 479–502.

Nakayama, A. (2019). Personal Communication.

Thomson, R. I. (2012). English Accent Coach [Computer program]. Version 2.3 www.englishaccentcoach.com

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference* (pp. 88-89), Ames, IA: Iowa State University.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*, 35–44.

Werker, J. F., & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63.

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, *52*, 349–355.

VI. Appendices

Appendix A: Recruitment Script

Researcher: Hello, I'm doing a study that compares the benefits of different methods to improve English pronunciation. Are you interested in participating?

<If the potential participant indicates their interest, I will give them a copy of my consent form.>

Researcher: Is it ok if I email you to schedule a time for us to meet and do the exercises?

<Participant answers yes, or we discuss their preferred method of contact.

Ok, Thank you. Have a nice day!

Appendix B: Informed Consent Form

Dear participant:

You are invited to take part in a research project that is part of my requirements for the Honors College at The University of Mississippi. This research project will be supervised by my thesis advisor, Dr. Vance Schaefer.

The purpose of the research is to determine whether listening to the sounds of English can improve pronunciation, or if practice producing those sounds is also necessary.

If you take part in my research, you will be recorded pronouncing words and do such activities as practice listening to words and choosing the sounds you hear, or reading words from a list.
You will come to the computer lab in Bondurant C-006 on three separate days in the same week.

Recordings and data collected will *not* be able to readily identify you, either directly or indirectly, and all of the recordings and data will be stored on an encrypted device.

Risks:
You may feel some performance related stress while being asked to identify sounds in words you hear or while pronouncing words. You may become tired or bored from sitting in front of a computer pressing the same one or two buttons for 20 minutes at a time.

Benefits:
You will get to practice listening carefully and identifying sounds in English, and you may experience satisfaction from contributing to scientific knowledge.

You are free to quit this research at any time. If you have any questions or concerns, please email me at ccrosby1@go.olemiss.edu. Thank you for your help.

Sincerely,

	Caleb Crosby	Vance Schaefer C-115 Bondurant Hall, P.O. Box 1848 University, MS 38677-1848
Phone	601-695-4706	662-915-1194
Email	ccrosby1@go.olemiss.edu	schaefer@olemiss.edu

IRB Approval

This study has been reviewed by The University of Mississippi's Institutional Review Board (IRB). If you have any questions, concerns, or reports regarding your rights as a participant of research, please contact the IRB at (662) 915-7482 or irb@olemiss.edu.

Statement of Consent

I have read and understand the above information. By completing the survey/interview I consent to participate in the study.

□ *By checking this box, I certify that I am 18 years of age or older.*

Signature: ______________________________ Date: _________________

Email: __________________________________

Appendix C: Personal Communication with Ai Nakayama

2019.09.25 Wednesday
20:27 Caleb Crosby What types of pronunciation problems cause the most difficulty in understanding what students are saying?
21:15 Ai Nakayama 中山藍 For Japanese students learning English, r/l confusions definitely cause problems.
21:16 Ai Nakayama 中山藍 Also b/v
21:16 Ai Nakayama 中山藍 F/h
21:16 Ai Nakayama 中山藍 Absolutely consonant clusters

Appendix D: Pretest vs Posttest Scores and Average Score by Individual Test Item

Participant		Pretest		Posttest	
Number	Group	word	Rating	word	Rating
2	A Grad Student	Van	3	Bad	3
		Best	3	Van	3
		Vine	1	Bush	3
		Fear	2	Book	3
		House	3	Vote	2
		Help	3	Vast	3
		Feed	3	First	3
		Vast	3	Have	3
		Form	2	Hut	3
		Bead	3	Had	3
		Have	3	Fall	3
		Bob	3	Feed	3
		(average)	2.66 (1.91)	(average)	2.91 (2)
3	A IEP student	Van	1	Bad	3
		Best	3	Van	3
		Vine	1	Bush	3
		Fear	3	Book	3
		House	3	Vote	1
		Help	3	Vast	2
		Feed	3	First	2
		Vast	3	Have	3

		Form	3	Hut	3
		Bead	2	Had	3
		Have	3	Fall	3
		Bob	3	Feed	3
		(average)	2.58 (1.83)	(average)	2.66 (1.91)
6	B Grad Student	Van	2	Bad	3
		Best	3	Van	3
		Vine	3	Bush	3
		Fear	2	Book	3
		House	3	Vote	3
		Help	3	Vast	2
		Feed	3	First	3
		Vast	3	Have	3
		Form	1	Hut	3
		Bead	3	Had	3
		Have	3	Fall	3
		Bob	3	Feed	3
		(average)	2.66 (1.91)	(average)	2.91 (2)

Numbers in parentheses show what the averages would be if 3-point rated words were collapsed into the 2-point category.