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*, ISSN 2380-9566, Dallas, TX, October 2015 (pp. 88-97). Ames, IA: Iowa State University.

IS PHONEMIC TRAINING USING NONSENSE OR REAL WORDS MORE EFFECTIVE?

Ron I. Thomson, Brock University
Tracey M. Derwing, University of Alberta & Simon Fraser University.

In this exploratory study we investigate whether focusing learners' attention on phonetic form through the presentation of vowels in isolated open syllables leads to greater learning than presentation of vowels in real words. Thirty-one intermediate, mixed-L1, English learners were assigned to three experimental conditions: A phonetically-oriented group (n=9); a word-focused group (n=12) and a control group (n=10). Using a high variability phonetic training (HVPT) paradigm, learners were taught to identify ten English vowels. Randomized recordings of the learners' pronunciation before and after training were evaluated by expert judges. Results indicated that the phonetically-oriented training was superior in promoting improvement in the pronunciation of real words relative to training almost entirely focused on the pronunciation of those same words. Furthermore, we found evidence to suggest that the accuracy of pronunciation before and after training was affected by interactions between the lexical frequency of the word in which the vowel appeared and each vowel's surrounding phonetic environment. Finally, implications for teaching are briefly addressed.

INTRODUCTION

The past decade has witnessed a notable increase in second language (L2) pronunciation research. Furthermore, many recent studies examining the efficacy of pronunciation instruction for adult learners indicate that instruction often makes a positive difference (Lee, Jang & Plonsky, 2015). However, although increased attention to what was previously an underrepresented area of inquiry is heartening, many gaps in the breadth and quality of the emerging research remain (Thomson & Derwing, 2015). Specifically, much of the current literature focuses on the efficacy of traditional approaches to teaching pronunciation, or repeatedly investigates the same questions rather than seeking new knowledge. For instance, the number of studies investigating English /l/-/r/ acquisition by Japanese speakers are added to regularly, but with little new insight gained (Thomson, 2011). Thomson and Derwing (2015) argue that the means of assessing development in pronunciation studies should be addressed. They particularly critique the over-reliance on reading tasks to test pronunciation. While easy to administer, these tasks do not provide an ideal representation of learners' speech in the context of real world communication. In this exploratory study, we attempt to fill gaps in the types of questions asked, and in the methods used to assess performance.

Background

It is widely accepted that most adult L2 learners do not develop native-like pronunciation of L2 vowels and consonants (Abrahamsson & Hyltenstam, 2009; Baker & Trofimovich, 2006; Bongaerts, van Summeren, Planken, and Schils, 1997; Munro & Derwing, 2008; Munro, Derwing, & Thomson, 2015). This reality has led many researchers and practitioners to search for the holy grail of pronunciation instruction: a method that can overcome this seemingly insurmountable obstacle. We do not view this search as productive, but instead agree with Levis (2005), who argues that approaches that treat accent-free L2 speech as the goal are both unrealistic and unnecessary. When it comes to the teaching of segmentals, our objective is to help learners better recognize and produce L2 speech sounds in ways that allow variation of pronunciation within individual categories. This approach accepts that how any given speech sound is produced can vary substantially without being misperceived by interlocutors as a member of a different category. Promoting within-category variation is in keeping with the fact that native speakers also differ in how they pronounce the same phonemes in different words or phonetic environments – what phonologists term allophonic variation. One obvious example is the difference in how English /l/ is pronounced in the word 'like' compared to the word 'ball'. Such allophonic variation is not always as obvious, however. In fact, all speech sounds are influenced by their neighboring phonetic environments. Consider the pronunciation of /u/ in 'boo' and /u/ in 'goo.' In the former word the vowel is produced much further forward in the vocal tract than in the latter.

Accepting that within-category variation is natural suggests that teaching the pronunciation of L2 sounds should incorporate and emphasize variation rather than focusing on elusive prototypes, citation forms, and the pronunciation of sounds in isolation. One technique that allows for a controlled approach to input variability is High Variability Phonetic Training (HVPT). Based on laboratory studies by Logan, Lively & Pisoni (1993), HVPT trains learners to identify L2 sounds in the context of stimuli spoken by multiple talkers and in multiple phonetic environments. Learners are first asked to indicate which speech sounds they perceive and are then provided with feedback on the accuracy of their responses. Manipulating training stimuli in terms of the number of talkers or the number of phonetic contexts in which sounds are presented is simple, and allows for determining whether training extends to new talkers and contexts (see Thomson, 2011; 2012a).

Since HVPT provides learners with feedback on the accuracy of their perceptions, it can help to direct their attention to properties of segmental stimuli important for L2 category formation. This is valuable because noticing linguistic forms in natural speech is often challenging for adult learners, who tend to focus on meaning instead of form (Schmidt, 2001). Guion and Pederson (2007) conducted an experiment in which adult L2 learners were trained to perceive L2 Hindi contrasts using identical stimuli, but one group was tasked with learning the meaning of words that differed by a single sound, while another group was asked to pay attention to the sounds of stimuli and how they differed, without knowing their meaning. The sound-oriented group performed significantly better on a discrimination test after training than the group that had been asked to focus on learning the meanings of the words.

The present study extends previous research by examining whether phonetically-oriented HVPT

training or lexically-oriented HVPT training has a greater impact on L2 learners' pronunciation of English vowels.

Research Questions

The following three research questions guided our study:

1. Does perceptual training using nonsense words *or* training predominantly focused on real words result in better pronunciation of real words?

2. Does degree of pronunciation improvement depend on the type of assessment task used?

3. Does pronunciation accuracy differ across English vowel categories and phonetic/lexical contexts?

METHODS

Participants

Thirty-six adult immigrants studying fulltime in a Canadian English as a Second Language (ESL) program volunteered. All were selected on the basis of having similarly assessed English abilities (LINC 5 – which is intermediate proficiency). Thirteen were randomly assigned to each of two experimental groups, and ten to a control group. Immediately after commencing the study, four participants withdrew from the first experimental group and one from the other, citing a lack of time to devote to the training portion. This left 31 participants, whose details are provided in Table 1.

Table 1.
Demographics of English learner participants

	Phonetic Group (n=9)	**Real Word Group (n=12)**	**Control Group (n=10)**
Age	M=37.3 (range: 24-46)	M=36.4 (range: 25-48)	M=33.1 (range: 23-40)
Sex	7 female; 2 male	9 female: 3 male	7 female; 3 male
Length of Residence	M= 12.7 months (range: 5-27)	M=20 months (range = 4-96)	M=15.5 months (range 5-57)
L1	Chinese (6); Amharic (2); Spanish (1)	Chinese (4); Spanish (2); Arabic (1); Russian (1); French (1); Romanian (1); Kinyarwandan (1); Punjabi (1)	Chinese (1); Spanish (2); Amharic (1); Arabic (2); Russian (1); French (1); Igbo (1); Zo (1)

On the surface it appears that the three groups are different in terms of their mean Length of Residence (LOR). In fact, these differences are largely attributable to a single outlier in the Real Word and Control groups, as reflected in the LOR range. When the single most extreme LOR outliers are removed from each group, their mean LORs are 13.6 and 10.9 months respectively. It should also be noted that the LOR outliers reported having very little interaction in English outside of the classroom. This, combined with their similar proficiency level, provides some assurance that LOR will not be an important factor in this study.

Perceptual Training

Using English Accent Coach (EAC) (Thomson, 2012b), learners in both experimental groups were trained to better recognize ten English vowels /i, ɪ, e, ɛ, æ, ɑ, ʌ, o, ʊ, u/. This freely available web-based HVPT application (www.englishaccentcoach.com) can be used to present to learners with isolated open syllables (i.e., just a consonant + vowel) or words containing target vowels and consonants, spoken by thirty speakers of Canadian English (similar to General American). Learners must respond to each item by clicking on the phonetic symbol associated with the vowel or consonant they believe each item contains (e.g., the identity of the initial consonant, stressed vowel, final consonant, etc.). After making their choice, they receive auditory and visual feedback on the accuracy of their selection. In this study, a researcher mode of EAC was used to precisely control the stimuli presented to each of the two experimental groups.

The first experimental group (Phonetic Group) received perceptual training for English vowels in the context of isolated open syllables (e.g., /bi/, /pi/, bɪ/ and /pɪ/); only 25% of their training sessions incorporated 70 target words (seven words containing each of the ten vowels). In the case of isolated CV syllables, many resulting tokens are not real words (e.g., /hɪ/, /hɛ/, /hʊ/), thus forcing learners to attend to phonetic information to successfully identify the vowels.

The second experimental group (Real Word Group) was trained almost entirely using the 70 target words. This group received only three brief phonetically-oriented sessions at the outset to ensure that they had learned the phonetic symbols, and a single phonetic session at the end. In CVC or more complex real words, learners are often able to recognize the word, but having recognized it, they may then apply knowledge concerning which vowel is supposed to occur in that word, as opposed to focusing on its phonetic properties.

Participants in both the Phonetic and Real Word groups completed 40 training sessions, at their leisure, over the course of one month, but were told they could complete a maximum of two sessions on any given day. Training sessions 1-3 comprised 100 items each, sessions 4-39, 150 items, and session 40, 200 items. The Control Group received no perceptual training. None of the groups received explicit articulatory training or practice. Like Thomson (2011, 2012a), we assume that improvement in perception will lead to changes in productions without any explicit pronunciation practice. Details concerning the training sessions are provided in Table 2.

Table 2

Perceptual training sessions by group and stimuli

Session	Phonetic Group	Real Word Group
1 – 3	Phonetic training: h+V syllables	Phonetic training: h+V syllables
4 – 29	Phonetic training: Open syllables presented in sets based on related consonant onsets. e.g., Session 1: p+V and b+V; Session 2: g+V and k+V, etc.	Target word training
30 – 39	Target word training	
40	Phonetic training: h+V syllables	Phonetic training: h+V syllables

Production Recordings

Before and after training, two tasks were used to elicit participants' productions of the same 70 target words used in training. In the first, participants heard the target words embedded in the carrier phrase, "The next word is ___," and they responded by repeating the word they had just heard in the carrier phrase, "Now I say ___." In the second task, a twenty-word subset of the target words (two nouns for each of the ten target vowels) were presented in the form of pictures. The participants created a sentence using each word. All tests were administered to participants individually in a quiet room and recorded using a high quality digital recorder and microphone. Individual productions of each target word from both the elicited imitation and picture-naming tasks were later extracted from the long recordings and saved as 5573 separate sound files representing individual productions of individual words (seven tokens were lost due to recording errors).

Judgments of Vowel Intelligibility

The 5573 individual sound files comprising participants' productions before and after training were presented to two phonetically trained judges (the authors) for evaluation using Praat's (Boersma & Weenink, 2016) Multiple Forced Choice script (see Thomson, 2013 for a detailed description). Files were blocked by each of the ten target English vowels, but randomly across speakers, time, and speaking task. These blocks of approximately 550 words each were further subdivided into five sets so that the task was more manageable and could be spread over a number

of sessions and days to avoid fatigue. Judges were asked to respond to each item by indicating whether it was a 'good' or 'poor' exemplar of the intended category, or another category altogether, assigning values of 2, 1, and 0 respectively.

RESULTS

We first examined the extent of agreement between judges. In 71% of cases, both were in complete agreement across the three possible categories (i.e., 'good', 'poor' or 'other'). When the good and poor examples were collapsed into a single category, agreement was at 84%. Given the borderline nature of some productions, where it was practically a coin toss between a very poor rendition of a particular vowel, or a poor rendition of the neighbouring category, this degree of rater agreement is extremely good. Because the judges' responses showed strong agreement, they were averaged for each item.

Three Bonferroni-adjusted Wilcoxon Signed Rank Tests were conducted to measure changes in performance over time on the elicited imitation task for the three groups. These indicated a statistically significant improvement in English vowel pronunciation for the phonetically-trained group (z = -2.695, p = .007) with a small effect size (r = .08). No significant improvement was detected for either the lexically-focused group or the control group.

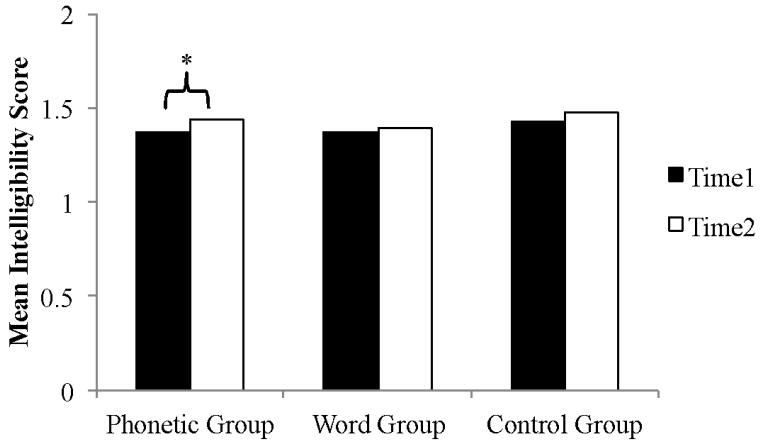

Figure 1. Mean intelligibility scores over time x group. The asterisk indicates a significant difference.

Three Bonferroni adjusted Wilcoxon Signed Rank Tests were also conducted to measure the groups' changes in performance over time on the picture naming production task (i.e., learners used target words in their own sentences). These tests indicated no significant difference for any group, although descriptively, the phonetic group demonstrated a larger improvement in the mean than the other two groups.

We conducted post-hoc analyses to examine whether improvement in the elicited imitation task was limited to particular vowels, or extended across all vowel categories. Learners in the

Phonetic Group demonstrated mean improvement in 8/10 vowel categories, /i, ɪ, e, ɛ, æ, o, ʊ, u/, with means scores for the remaining two, /ɑ/ and /ʌ/, decreasing between Time 1 and Time 2. Learners in the Real Word Group demonstrated mean improvement in 4/10 vowel categories /e, æ, ɑ, o/), with mean scores for the remaining vowels staying unchanged or decreasing. Finally, learners in the Control Group demonstrated mean improvement in 6/10 vowels /ɪ, e, ɛ, æ, ɑ, o/, with mean scores for the remaining vowels each decreasing. In sum, while scores on /i, ʊ, u/ increased for the Phonetic Group, they did not improve for the Real Word and Control Groups. Conversely, the mean score for /ɑ/ decreased for the Phonetic Group, while improving for the other groups.

We next examined individual trajectories in each group, finding that 89% of the Phonetic Group improved over time, with only one participant showing no improvement, but also no decline. In contrast, only 50% of the Real Word Group and 60% of the Control Group demonstrated improvement in their mean scores. There was no correlation between LOR and extent of improvement.

Finally, we examined whether particular L2 vowel categories were more intelligible than others, and whether the word in which they occurred played a role. Recall that each vowel occurred in seven distinct lexical contexts. Results indicated that four vowels /i, e, ɑ, o/ were more accurately produced than the remaining vowels, and that their scores were also least affected by the words in which they occurred. For example, the /i/ in 'bead' was as intelligible as the same vowel in 'leaf'. In contrast, the intelligibility of /ɪ, ɛ, æ, ʌ, ʊ, u/ varied dramatically, depending on the words in which they were produced. For example the /ɪ/ in 'bid' was rarely intelligible, while the same vowel in 'which' and 'stick' was usually intelligible. Similarly, the /ɛ/ in 'head' and 'jet' were very intelligible, while in 'bread' it was not. While word familiarity and frequency may play a role (Thomson & Isaacs, 2009; Munro & Derwing 2008) this does not easily account for much of the variation in the current data. For instance, while the vowel in 'bread' (high frequency) had a low score both before and after training, the same vowel in 'sketch' (low frequency) had a high score at both times. Furthermore, there is no clear evidence that the phonetic environment plays a primary role. For example, the vowel in 'bud' was far less intelligible than the same vowel in 'scum.' The latter has a complex onset, which does not appear to cause difficulty for speakers in producing the following vowel. Similarly, the vowels in 'spin' and 'stick' were both far more intelligible than in the word 'bid'.

DISCUSSION

This exploratory study provides preliminary evidence in response to our three research questions. First, does perceptual training using nonsense words *or* training predominantly focused on real words result in better pronunciation of real words? The findings suggest that at least for participants in this study, forcing learners to attend to phonetic details during perceptual training resulted in significant improvement in pronunciation. However, in answer to our second question, regarding whether improvement depends on the type of assessment task used, we found that detectible improvement is limited to more controlled productions. That is, it did not seem to transfer to more extemporaneous pronunciation of the same words. It is encouraging to see improvement in the elicited imitation task, however, since this task is more challenging than the reading tasks commonly used in this line of research.

Finally, we asked whether pronunciation accuracy differs across English vowel categories and across phonetic/lexical contexts. We found that, in general, English lax vowels, /ɪ, ɛ, æ, ʌ, ʊ/ as well as /u/ were the most challenging for learners, while the remaining four tense vowels /i, e, ɑ, o/ were relatively clear, perhaps because similar vowels appear in the learners' L1s. This seemed to be the case regardless of the phonetic or lexical context in which the vowels were found. While there may be some evidence that lexical frequency affects the intelligibility of vowel production, it is not a straightforward predictor. Furthermore, there is little evidence to support a general pattern with respect to vowel intelligibility being affected by the complexity of the surrounding phonetic environment. Therefore, it is more reasonable to conclude that a complex interplay among lexical frequency, phonetic context, and other factors accounts for differences in intelligibility scores for the same vowel category produced in different words. In some cases, a vowel found in a complex phonetic environment may be in a word that is so frequent that the learner had already acquired the vowel in that word.

IMPLICATIONS

Our findings suggest that instruction of L2 vowels should include some focus on phonetic level information, as opposed to focusing solely on the pronunciation of sounds in real words. This approach may help to draw learners' attention to phonetic information in a way that using only words as training stimuli does not. At the same time, the improvement demonstrated by the Phonetic Group in this study had a small effect size, while the Real Word Group showed no improvement after a significant amount of input. These marginal gains suggest that perceptual training on its own is insufficient to promote maximal improvement. Instead, learners likely need explicit practice producing the sounds they are learning to more accurately perceive. While the present study lasted a month, the hours of training was relatively small (approximately 10), and even smaller on a per vowel basis (1 hour). Thus, we speculate that longer training will have a stronger impact.

Finally, this study has implications for focus of training. As Munro and Derwing (2008) and Munro, Derwing & Thomson (2015) found, some English sounds appear to be easy for learners, in some instances, because there are direct parallels in their L1, and in other cases, they may simply be easy to perceive and produce. Knowledge of which categories are most challenging can allow teachers to focus on those sounds that are less amenable to natural improvement.

ACKNOWLEDGEMENTS

The authors thank Murray Munro for his input on study design. We are grateful to participating students, teachers and staff at NorQuest College, and to our research assistants: Isabel Light, Lynn Sawyer, and Jennifer Foote. We acknowledge the Social Sciences and Humanities Research Council of Canada for funding this research.

ABOUT THE AUTHORS

Ron Thomson is an Associate Professor of Applied Linguistics at Brock University, in Niagara, Canada. He teaches in the undergraduate program in Applied Linguistics, and in the MA TESL

program. His research interests include oral fluency and pronunciation development, oral fluency instruction, high variability phonetic training, and computer assisted pronunciation teaching.

Tracey Derwing is an Emeritus Professor in the TESL program at the University of Alberta, and an Adjunct Professor in the Linguistics Department at Simon Fraser University. She is interested in identifying ways to enhance communication across L1 and L2 learners.

REFERENCES

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning, 59*, 249–306.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching, 44*, 231-250.

Boersma, Paul & Weenink, David (2016). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.14, retrieved 11 February 2016 from http://www.praat.org/

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition, 19*, 447-465.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Second-language speech learning: The role of language experience in speech perception and production: A festschrift in honour of James E. Flege* (pp. 57-77). Amsterdam: John Benjamins.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics* 36, 345–66.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*, 369-377.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1993). Training listeners to perceive novel phonetic categories: How do we know what is learned? *Journal of the Acoustical Society of America, 94*, 1148-1151.

Munro, M. J., & Derwing, T. M. (2008). A developmental study of Mandarin and Russian speakers' English vowels. *Language Learning, 58*, 479-502.

Munro, M. J., Derwing, T. M., & Thomson, R. I. (2015). Implications of naturalistic segment development for pronunciation teaching technology. *International Review of Applied Linguistics in Language Teaching, 53*, 39-60.

Schmidt, R. (2001). Attention. In P. Robinson. (Ed.). *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Thomson, R. I. (2013). Phonology: Quantitative methods. In C. A. Chapelle. (Ed). *The Encyclopedia of Applied Linguistics.* (pp. 4434-4444). Oxford: Wiley Blackwell. DOI: 10.1002/9781405198431.wbeal0912

Thomson, R. I. (2012a). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning, 62,* 1231-1258.

Thomson, R. I. (2012b). *English Accent Coach* [Computer program]. Version 2.3

Thomson, R. I. (2011). Computer Assisted Pronunciation Training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, *28*, 744-765.

Thomson, R. I. & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*(3): 326-344.

Thomson, R. I., & Isaacs, T. (2009). Within-category variation in L2 English vowel learning. *Canadian Acoustics, 37,* (3), 138-139. Proceedings of the annual conference of the Canadian Acoustics Association.