A Report on the Design and Outcomes of a Perceptual Training Program Targeting English /r/-/l/

Leah GILNER

要　旨

本研究は，英語の /l/ および /r/ 音素に関する視聴覚知覚トレーニングプログラムの説明と，プリテストとポストテスト「処置前―処置後」データの結果を提供する。本研究の主な目的は，短期間ではあるものの，定期的に実施される多様な音声変異を含む音韻トレーニングが，オンラインプラットフォームを通じて独立して行われることで，学生に有益かどうかを確認する。事前・事後テストデザインを用いて，識別精度の全体的な向上を測定する。9 回のトレーニングセッション後に短いテストが実施され，プログラム終了時には学生の印象を探る自由回答式のアンケートが行われる。定量的なテスト結果からは，トレーニングプログラムの進行に伴い識別精度が向上したことが示され，定性的なアンケートの回答からは，学生が視覚的手がかりや，対象音素の知覚と産出に関連する調音動作の重要性に気付いたことが示唆される。

キーワード：高変動音素訓練, 視聴覚知覚トレーニング, 発音, 知覚, EIL

Keywords: High variability pronunciation training, audiovisual training, pronunciation, perception, EIL

1. Introduction

The audiovisual perceptual training program discussed in this study consists of a streamlined and consolidated version of the author's classroom practices, refined through years of experimental observation and reflective practice. The desire to make the training more widely accessible prompted the creation of video-based materials that could potentially be administered for autonomous self-study via a learning management system. This report provides a description of the training program in English /l/ and /r/[1] phonemes along with results from preliminary assessment tests.

The primary objective of this pilot study was to assess whether students would benefit from brief but regular, high-variability pronunciation training administered independently through an online platform. A pre- and post-test design was used to measure overall gains in identification accuracy. Short tests were conducted after each of the nine training sessions and an open-ended questionnaire probing students' impressions was given at the end of the program. The quantitative testing results indicate that identification accuracy improved over the course of the training program while the qualitative questionnaire responses suggest that students noticed the relevance of visual cues and associated articulatory gestures in both perception and production of the target phonemes.

The instructional approach combines articulatory phonetics with high-variability phonetic training. It draws attention to the muscle movements, or articulatory gestures, underlying speech production with particular emphasis on the role of the lips, tongue, and jaw. Multiple voices and a plurality of accents are used to exemplify the variability inherent to language production. Thus, the approach goes some way toward closing the gap between the narrow range of exposure provided by commercial instructional materials and the expansive reality of contemporary English language users who are these days from all parts of the globe. It can therefore be thought of as a kind of anticipatory preparation for using English as a tool for international communication.

While the program design can be implemented with any target phonemes, this pilot study focused on the lateral approximant /l/ and the approximant /r/. Researchers have long been interested in how Japanese L1 speakers process these sounds (e.g., Lively et al., 1993; Logan et al., 1991; Strange & Dittmann, 1984). There seems to be consensus that the crucial acoustic cue for distinguishing them is the F3 format frequency, which

is relatively higher for the lateral approximant /l/ than for the approximant /r/. The F3 frequency for the apico-alveolar tap /ɾ/ used in Japanese falls in between these two (Shinohara & Iverson, 2021). The articulatory correlates to the F3 format are primarily associated with the position and shape of the tongue, particularly in the palatal region of the oral cavity which influences the space between the tongue and the roof of the mouth, as well as lip rounding.

Audiovisual training has been shown to be an effective way to improve identification accuracy of English /l/ and /r/ among Japanese L1 listeners. Hardison (2005), for example, compared audio-only (A) and audio-visual (AV) perceptual training for words beginning with /l/ and /r/, among other sounds. The study involved university-aged English learners from Japan and Korea enrolled in an intensive program in the United States. A pre- and post-test design was used to track changes in identification accuracy after fifteen 45-minute sessions including perceptual training. The perceptual training consisted of video recordings of minimal pair words spoken by American English speakers. Findings indicate that AV input facilitated earlier identification of spoken words compared to A-only input. Training also led to significant improvements in word identification accuracy, particularly for words beginning with /l/ and /r/ and particularly for Japanese learners. These findings support the priming role of visual cues in AV speech processing and the effectiveness of perceptual training in improving L2 spoken word identification.

Findings from Shinohara (2021) provide further support for an AV advantage in perceptual training. This study involved Japanese L1 speakers ranging in age from seven to 23 who underwent ten 40-minute audiovisual training sessions consisting of video-recordings of minimal pair words spoken by Standard Southern British English speakers. Pre- and post-tests were administered under three different conditions: audiovisual (AV), visual only (V), and audio only (A). Interestingly, participants were told to look at the speaker's mouth movements when the video clip was played. Results showed that identification accuracy improved in all three conditions. Improvement was greatest in the AV condition, followed by the visual-only condition, then the audio-only condition. These findings are particularly noteworthy because listening as well as pronunciation instruction, if and when directly implemented, generally resemble the audio-only condition, which is to say, the condition found to be the least effective of the three.

Both of the studies just mentioned included high-variability phonetic training (HVPT) as part of the experimental design. HVPT refers to perceptual training that uses

numerous samples, produced by multiple speakers (usually of a single variety) in various phonetic contexts. Training samples commonly exemplify target sounds by means of minimal pair words. In a review of 32 HVPT studies provided by Thomson (2018, p. 214), it was observed that "97% resulted in significant improvement in learners' mean perception and/or production scores for trained L2 sounds". In addition, all of the studies demonstrated that this kind of training generalizes to new instances of the same sounds.

Given the empirical support for audiovisual training and the documented challenges faced by Japanese L1 learners in distinguishing /l/ and /r/ sounds, this study seeks to contribute further by exploring whether incorporating a wider variety of accents can enhance the robustness of phoneme identification in an autonomous learning environment.

The training program under discussion here implements a *pluralistic* HVPT design in an effort to expose students to comparatively more and varied speech patterns. Commercial instructional materials generally confine models to one particular variety and a limited number of speakers. For example, Levis (2016) noted that his graduate students undertook an analysis of commercial pronunciation software programs and found that most had only one speaker for all activities. He also noted the importance of conducting studies which explore the possibilities that exist for the range and type of multiple talkers. The approach presented here is pluralistic and expands on common HVPT practice by including several varietal accents from, namely, Australia, Hong Kong, India, Ireland, Nigeria, Philippines, Singapore, the United Kingdom, and the United States. The choice of these accents was based on the range and quality of voices available by means of online technologies.

This pilot study was devised with the aim of documenting the development and implementation of the program as well as measurement of its outcomes. It seeks to answer the following research questions:

RQ1. Can students benefit from an independent and extracurricular audiovisual training program in English /l/ and /r/ phonemes?
RQ2. Are there observable differences in results between groups who received different types of classroom implementations?
RQ3. How do students describe their understanding of English /l/ and /r/ after the training program?

Pre- and post-test performance data collected before and after the intervention along with weekly training test data will be used to answer the RQ1 and RQ2. Open-ended questionnaire responses will be used to answer the RQ3.

2. Methodology

While the program was initially designed for autonomous self-study, it was administered to three intact elective English courses at a Japanese university under supervision of the author and creator for the purposes of this study. Two of the courses focused on listening and speaking skills (Group A and Group B) and one on writing skills (Group W). The inclusion of Group W makes it possible to compare outcomes between groups receiving different types of classroom implementations. The most pertinent difference between the two classroom implementations was that course-related activities for Groups A and B included extensive listening to conversational EIL with speakers from Africa (Nigeria, South Africa), the Americas (Chile, Honduras, Mexico, US), Asia (China, Hong Kong, India, Indonesia, Mongolia, South Africa, Sri Lanka), Europe (Belgium, Croatia, Portugal, Spain), Oceania (Australia), and the UK (England)[2].

2.1. Training materials

The audiovisual training program is comprised of an introductory session and nine training sessions. The words included for testing (n=40) and training (n=45) were selected from those used in Iverson, Hazan, and Bannister (2005). Program sessions are presented by means of stand-alone videos and each one is approximately two minutes long as shown in Table 1. Audio samples were curated from the corpus of voice recordings freely available online. *Audacity* editing and recording software (https://www.audacityteam.org/) was used to prepare studio-quality recordings. Including the introductory session, the entire program totaled approximately 26 minutes.

The video materials, created with PowerPoint, have been designed to help students notice certain muscle movements associated with each phoneme. Appendix A provides the complete PowerPoint slides of the Introduction session in handout form. Making use of the analogy with sports and how each sport uses muscles differently, viewers are visually guided through a series of images illustrating iconic poses associated with various sports postures and then a parallel is drawn with sounds in language.

Table 1. Descriptive details of the training program

	Varieties	Length
Pre- / Post-test	UK	
Introduction	US	6:32
Training 1	Hong Kong, India, Ireland	2:24
Training 2	Australia, India, Nigeria	2:07
Training 3	Ireland, Philippine, Singapore	2:18
Training 4	Australia, Hong Kong, Singapore	2:18
Training 5	India, Nigeria, Philippine	2:21
Training 6	Australia, Ireland, Philippine	2:13
Training 7	Hong Kong, India, Singapore	2:19
Training 8	Australia, Hong Kong, Ireland	2:12
Training 9	Nigeria, Philippine, Singapore	2:18

Seeing speech (Lawson et al., 2018) animated productions are used to illustrate the articulatory gestures corresponding to the voiced alveolar tap shown with Japanese syllabary symbol, the voiced alveolar lateral approximant shown with the roman letter L, and the voiced alveolar approximant shown with the roman letter R. Each phoneme is associated with a particular color scheme, a colored rectangle draws attention to the position of the articulators, and articulation hints are provided. Next, each of the three phonemes is shown with a colored drawing of a frontal view of the mouth shape for each one designed for lip sync work (Mozart3737, n.d.). Then, the tap phoneme is eliminated and information for only the L and R phonemes are presented with the message:

different sound = different word

different word = different meaning

The Introduction closes with the presentation of 14 minimal pairs exemplifying word initial, intervocalic, and consonant cluster positions. On the left side of the screen, the visual cues and articulation hints for the L phoneme are shown and the L words are heard in isolation spoken by a female and male voice in a UK accent. Then a blank screen is shown and the information for the R phoneme is provided on the right side of the screen. The R words are heard in isolation spoken by the same female and male voices. Finally, the two preceding presentations are integrated, showing L information on the left and R information on the right while each L and R minimal pair are heard spoken by the same female and male voices. The introduction video concludes by replacing one of the

visual cues with the message:

different sound = different word

different word = different meaning

Each training session targeted six minimal pairs spoken by a male and female voice in each of three different varietal accents. The nine training sessions integrated the color scheme, visual cues, and articulatory hints used in the Introduction into the presentation format. Training sessions 1 through 5 included all three types of information for each phoneme. The forced-identification task administered after these sessions required choosing between two choices after hearing an audio prompt. The audio prompts were drawn from the training samples. Figure 1 provides screenshots of the training session presentation along with the testing task format.

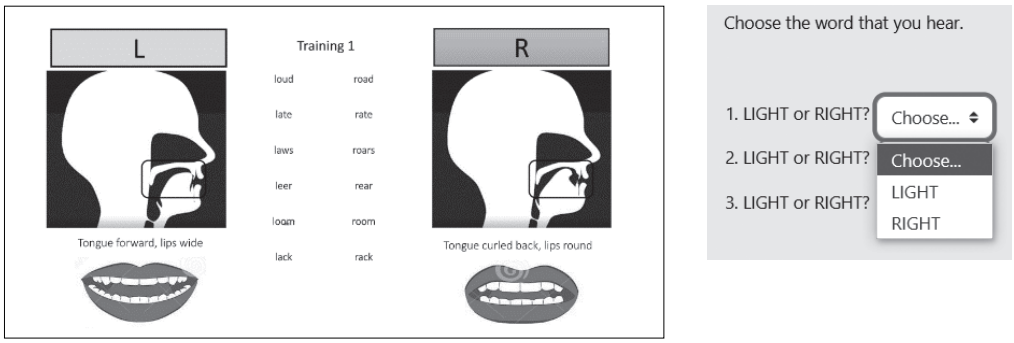


Figure 1. Training session (left) and testing presentation (right) format (sessions 1 through 5)

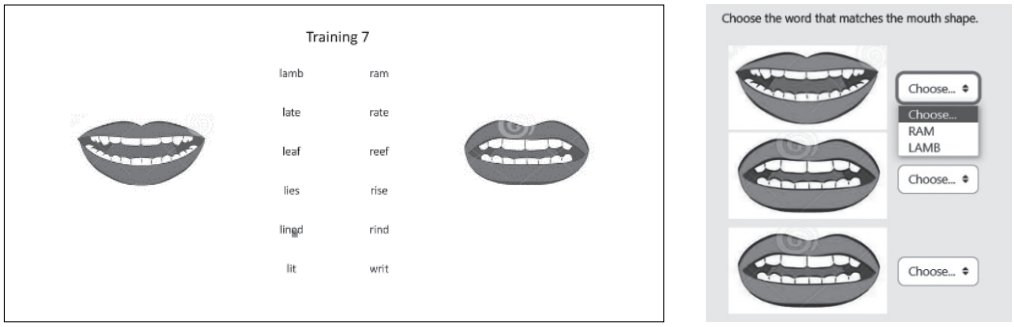


Figure 2. Training session and testing presentation format (sessions 6 through 9)

Training sessions 6 through 9 focused on visual cues associated with each phoneme. The forced-identification task administered after these sessions required choosing the word that matched the visual cue provided, as shown in Figure 2.

3. Results

3.1. Pre-test and Post-test

Changes in pre- and post-test scores indicate improvement in identification accuracy in all three groups after the nine weeks over which the training program was administered. The descriptive statistics provided in Table 2 show that average pre-test scores varied across the groups. A notable distinction between pre-test scores for the two groups enrolled in oral communication courses can be seen. The average score for Group A (62.85%) was 14.41% lower than that of Group B (77.26%). Group W's average pre-test score fell in between the other two groups at 66.58%. Groups A and B made similar gains over the course of the training as shown by the differences in pre- and post-test scores, 7.44% and 7.14%, respectively. Standard deviations (SD) indicate little change in score variability for Groups B and W. Group A shows higher variability in the post-test (16.54%) compared to the pre-test (14.27%) suggesting relatively less uniformity in performance.

The trendlines in Figure 3 reveal parallel improvement across groups, with Groups A and B, both focused on oral communication, exhibiting steeper inclines compared to Group W, which had a writing focus. This supports the hypothesis that phonetic training is more effective when combined with active listening and speaking practice, suggesting a classroom implementation effect.

A paired-samples t-test was conducted to compare pre-test and post-test scores. The results demonstrated a significant improvement in phoneme identification accuracy

Table 2. Descriptive statistics for Pre- and Post-test scores

	N	Pre-test average % / 40	Post-test average % / 40	Difference	Pre-test SD	Post-test SD
Group A	26	62.85% (25.14)	70.29% (28.12)	7.44%	14.27%	16.54%
Group B	21	77.26% (30.90)	84.40% (33.76)	7.14%	16.89%	16.94%
Group W	19	66.58% (26.63)	72.11% (28.84)	5.53%	16.51%	16.55%

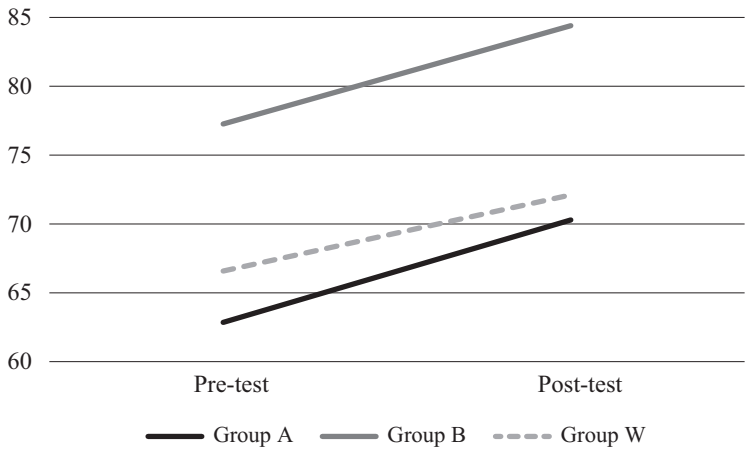


Figure 3. Pre- and Post-test trendlines for the three groups

across all groups: pre-test (M = 68.90, SD = 7.48) and post-test (M = 75.60, SD = 7.67), t(2) = −11.30, p = 0.007 (two-tailed). The effect size (Cohen's d = 0.88) indicates a large practical significance, further suggesting that the training not only led to statistically significant improvements but also had a considerable impact on students' ability to distinguish between the target phonemes.

Summing up, all three groups showed improvement in their post-test scores, indicating that the training was generally effective. At the same time, the two groups enrolled in oral communication courses showed greater improvement than the group enrolled in the writing course. Group A displayed an increase in variability, which may indicate a more diverse response to the training. This could be due to the relatively larger proportion of first-year students in Group A (76%) compared to the other two groups (Group B = 52%, Group W = 0).

3.2. Training tests

Table 3 presents the average test scores of the three groups after each training session (T1 to T9). It is apparent that values increased over time, indicating improvement in identification accuracy. Group W had the highest starting score at 12.95 in T1 and continued to improve up to T4 (14.58). A slight dip in performance is noted at T5 (12.63), suggesting variability in the learning process. Scores thereafter increased reaching 17.45 at T7 and 18.00 at T8, indicating substantial improvement. The final score at T9 was

17.67, showing consistent performance.

Table 3. Average test scores for the nine training sessions

	T1	T2	T3	T4	T5	T6	T7	T8	T9
Group A	11.71	11.33	11.33	11.77	12.86	16.04	16.43	17.54	17.78
Group B	11.27	12.11	11.89	12.81	12.63	15.78	17.95	18.00	17.75
Group W	12.95	13.79	13.92	14.58	12.63	14.77	17.45	18.00	17.67

Group A and Group B began with similar average scores, 11.71 and 11.27, respectively. Scores on T5 are also similar for these two groups, although Group A made relatively steadier progress and larger gains between T4 and T5, increasing from 11.77 to 12.86. From T1 to T2, the scores for Group B increased from 11.27 to 12.11. Another notable increase is observed between T3 (11.89) and T4 (12.81). Taken together, these scores suggest small positive changes in the early stages of the training.

Figure 4 illustrates the changes over the course of the nine sessions. From T5 to T6, scores for all groups increased notably (Group A = 17.67%, Group B = 17.50%, Group W = 11.89%) and remained above 90% for the rest of the sessions. It is relevant to note that these high scores were obtained on the visual-cue matching tests.

3.3. Open-ended question responses

A total of 40 responses to an open-ended question were collected from students in Group A and Group B. The precise wording of the question was as follows:

What did you learn from L and R training? [more than 50 words]

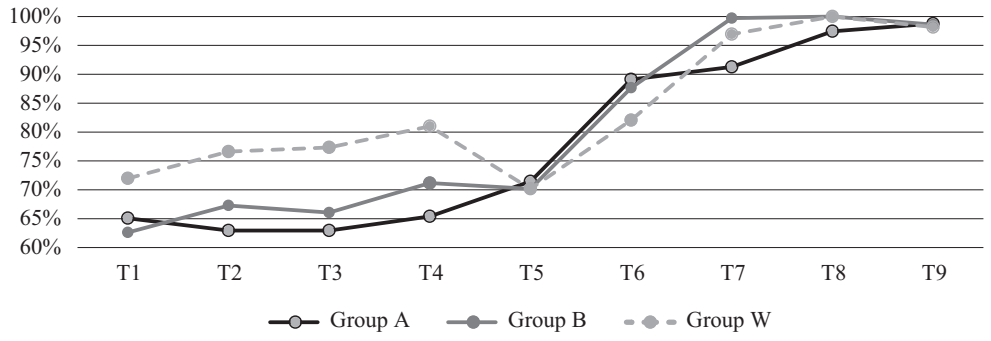


Figure 4. Test results for nine training sessions

Lexical and thematic analyses were carried out using KH Coder and Microsoft Excel. The responses were on average 54 words and taken together totaled 2,164 running words. Figure 5 displays the words that occurred 10 times or more in the collection of responses along with visualizations for the two most influential clusters from a hierarchical cluster analysis created with KH Coder.

The visualizations show how these words are related to one another based on their co-occurrence in the dataset. The bar chart to the left of the words highlights

Word	Occurrences
pronunciation	64
pronounce	
pronounced	
difference	55
different	
differentiate	
differently	
training	29
understand	26
learn	24
mouth	23
word	22
able	20
ability	
distinguish	17
listen	17
think	17
shape	16
sound	15
improve	12
tongue	12
class	11

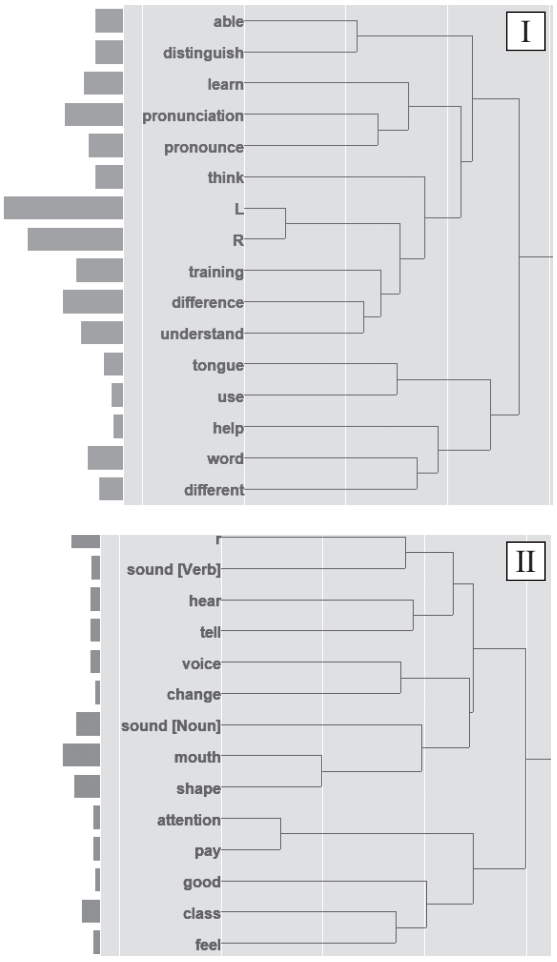


Figure 5. Most frequently occurring words in the open-ended question responses and their hierarchical relationships

the frequency of each term. For instance, terms such as “pronunciation,” “learn,” and “training” have relatively larger bars, indicating that these words occur more frequently in the responses. Terms such as “different” and “word” have shorter bars, suggesting lower frequency. The words that are more closely related (based on their context or usage) are connected by short vertical lines. Words that are less similar are connected higher up in the hierarchy with longer branches.

The hierarchical structure of Cluster I shows how certain words are conceptually or contextually closer to one another in the dataset, with a clear progression from core concepts like “pronunciation” and “learn” to more specific terms like “tongue” or “help.” We can see that words like “pronunciation,” “learn,” “pronounce,” “L,” “R,” and “think” are closely clustered together, indicating a strong relationship between these words based on their frequent co-occurrence in the collection of responses. Words higher up in the tree, like “able” and “distinguish,” are less connected to the central cluster but still form meaningful connections with the more central words. Words lower in the hierarchy, such as “help,” “use,” and “tongue,” are connected through broader relationships.

The central terms in Cluster II are “sound [Noun]”, “mouth”, “shape”, “voice”, and “change”. The presence of both “sound [Verb]” and “sound [Noun]” at the top of the hierarchy suggests a strong thematic focus on sound in this part of the data. “Sound” in its different forms (as a verb and a noun) may refer to both the production of sound (as in pronunciation or speaking) and the perception of sound (as in hearing and listening). Closely related terms like “hear,” “voice,” and “tell” further reinforce the idea that this cluster deals with the auditory aspect of communication—both producing and perceiving sound. The term “mouth” and its connection to “shape” suggest an emphasis on articulation and how the mouth, tongue, and vocal tract shape different sounds during speech production. This aligns with the broader focus on pronunciation and sound learning, where the physical manipulation of the mouth and shape are crucial to producing accurate sounds. The term “attention” appears in a separate branch, but it connects with terms like “pay” and “good,” which may suggest a cognitive aspect to the analysis. This could relate to the mental focus or cognitive effort required to pay attention to sound or speech. This cluster highlights the acoustic and articulatory aspects of learning, particularly focusing on sound production and perception.

This lexical clustering analysis aligns with prior research suggesting that learners often draw connections between their articulatory awareness and their phonemic

perception. Such awareness, particularly in distinguishing between /l/ and /r/, appears to be a critical factor in improving both identification and production.

Results from thematic analysis complement and supplement these lexical trends. Four prominent themes emerged from manual inspection of the collection of responses, as shown in Table 4. Generally, more than half of the responses included two themes (*n* = 28), nine responses included one theme, and three responses included three themes. Furthermore, more than half of the responses (*n* = 23) referred to learning outcomes related to both perception and production, nine referred to outcomes related only to production, and four to outcomes related only to perception.

Table 4. Prominent themes observed in the open-ended questions

Themes	Number of mentions
Perceptual distinctions	33
Articulation techniques	20
Mouth shapes	13
Speaker variability	6

Perceptual distinctions and *Articulation techniques* were found to be the most prominent themes. Among the 40 responses, 83% made reference to the ability to distinguish the two sounds and 50% referred to articulation and/or the articulators. Six responses spoke directly to a previous lack of 'knowing' there was a difference between the L and R. A similar number expressed having become 'better' at understanding differences between them. More than one-third (33%) of the responses included specific reference to the shape of the mouth and 15% referred to noticing differences in pronunciation among speakers. Table 5 presents representative examples for each theme taken from the dataset.

4. Discussion

With these results in mind, we now turn our attention back to the research questions formulated in the Introduction.

Table 5. Representative examples of each theme taken from the dataset

Themes	Examples from the dataset
Perceptual distinctions	"In the L and R training, I was able to understand the difference between raw and law, read and lead for example, and other detailed differences between each word."
	"Before taking this class, I couldn't distinguish between L and R very much. However, every time I take a class, I distinguish between L and R, so I understand the difference little by little."
	"It was sometimes difficult to distinguish between the L and R words by only listening, but through the L and R training, I found myself able to understand the differences in the sounds."
Articulation techniques	"When pronouncing the R, the tongue curls. Also, when pronouncing the L, put the tongue above the mouth."
	"L and R are different from tongue position, and I thought the training made me used to the difference."
	"The first thing I realized is that the 'L' sound comes from the front of the mouth, while the 'R' sound comes from the back of the mouth."
Mouth shapes	"I could completely discern the difference between the L and R mouth shapes now."
	"I learned that my ability to distinguish between L and R has improved since taking this class. I can now tell the difference between the mouth shapes of L and R."
	"In the case of L, the shape of the mouth spreads sideways. In the case of R, the shape of the mouth is a little vertically long compared with L."
Speaker variability	"Even men and women speak differently, so I want to be able to distinguish between them."
	"I learned that there are several ways to pronounce L and R words even though the words are the same."
	"…after all of the training i think i can understand why we need to do this. Actually many of the people form the different countries have different speaking habits."

4.1. RQ1. Can students benefit from an independent and extracurricular audiovisual training program in English /l/ and /r/ phonemes?

Comparison of pre- and post-test scores suggests that students can indeed benefit from an independent audiovisual training program in English /l/ and /r/ phonemes. All three groups showed significant improvement after completing training, thereby reinforcing previous findings on audiovisual HVPT (e.g., Hardison, 2005; Shinohara, 2021). The findings presented here also suggest that audiovisual HVPT which incorporates articulatory phonetics by emphasizing the visual cues and articulatory gestures of sounds could potentially be effectively applied to the creation of autonomous learning materials.

4.2. RQ2. Are there observable differences in results between groups who received different types of classroom implementations?

Notably, significant differences emerged between the groups, with those in oral communication courses showing greater improvements compared to the writing group. Group A, which had more first-year students, started with lower pre-test scores (62.85%) compared to Group B (77.26%) and Group W (66.58%). Despite the differences in initial performance, both Groups A and B, which focused on listening and speaking, made similar progress with approximately 7% improvement, while Group W, which focused on writing, improved by 5.53%. These differences suggest that the type of classroom focus, particularly one that included extensive listening to EIL conversations, may affect how much students benefit from phonetic training programs.

4.3. RQ3. How do students describe their understanding of English /l/ and /r/ after the training program?

Students described their improved understanding of /l/ and /r/ phonemes in terms of both perceptual and articulatory gains. Common themes in the open-ended responses included an increased ability to distinguish between the two sounds, recognition of differences in mouth shapes, and awareness of the role of the tongue's position. Many students emphasized the importance of articulatory gestures in distinguishing /l/ and /r/, and several noted that their ability to differentiate between these sounds improved over time. For example, students mentioned understanding differences such as "raw" vs. "law" and noticing how tongue positioning affected pronunciation.

5. Closing remarks

This pilot study was undertaken with the goal of ascertaining potential benefits of extracurricular and independent HVPT. All in all, the results are encouraging. They suggest that it is possible to produce effective audiovisual HVPT materials which expose students to not only multiple voices but also various accents using readily accessible technologies. It merits repeating that identification accuracy improved significantly after the training program under investigation here which is comprised of nine 2-minute training sessions and just under 30 minutes long in total. This is a remarkably shorter intervention than HVPT interventions tend to be. Recall that the treatment in Hardison (2005) was comprised of fifteen 45-minute session (almost 11 hours 30 minutes) and the one in Shinohara (2021) was comprised of ten 40-minute sessions (almost 7 hours). The approach implemented in this pilot study could thus help make perceptual pronunciation training more widely available to interested parties since high-quality, video-recorded materials may not be a viable option for many (as is the case for the author).

These findings hold promise for developing scalable phonetic training programs that can be used in diverse educational contexts. Future studies should explore the potential of extending this methodology to other phonetic contrasts and learner populations. A primary motivation for designing the program as well as carrying out this pilot study was to assess the viability of the training method. It was administered under supervision by the author (and creator) in an effort to confirm that it was completed as intended and thus establish a point of reference for subsequent administrations as a self-directed autonomous study activity. The video materials are available on ResearchGate in the event that others are interested in contributing to this line of research.

Notes

1 For convenience, the /r/ symbol is used throughout this discussion to represent the alveolar approximant. It should be noted that the corresponding IPA symbol is /ɹ/.

2 The conversations used for extensive listening were selected from the treasure trove of materials made freely available on English Listening Lesson Library Online (elllo.org). A big debt of gratitude is owed to Todd Beuckens for creating and maintaining this invaluable resource.

References

Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics*, *26*(4), 579–596. https://doi.org/10.1017/S0142716405050319

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 12.

Lawson, E., Stuart-Smith, J., Scobbie, J. M., & Nakai, S. (2018). *Seeing Speech: An articulatory web resource for the study of Phonetics*. University of Glasgow. https://www.seeingspeech.ac.uk/ipa-charts/?chart=3

Levis, J. (2016). Research into practice: How research appears in pronunciation teaching materials. *Language Teaching*, *49*(3), 423–437. https://doi.org/10.1017/S0261444816000045

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://doi.org/10.1121/1.408177

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. https://doi.org/10.1121/1.1894649

Mozart3737. (n.d.). *Mouth-sync-talking-lips* [Graphic]. dreamstime.com

Shinohara, Y. (2021). Audiovisual English /r/-/l/ Identification Training for Japanese-Speaking Adults and Children. *Journal of Speech Language and Hearing Research*, *64*, 2529–2538. https://doi.org/10.1044/2021_JSLHR-20-00506

Shinohara, Y., & Iverson, P. (2021). The effect of age on English /r/-/l/ perceptual training outcomes for Japanese speakers. *Journal of Phonetics*, *89*, 101108. https://doi.org/10.1016/j.wocn.2021.101108

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131–145. https://doi.org/10.3758/BF03202673

Thomson, R. I. (2018). High Variability [Pronunciation] Training (HVPT): 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

Appendix

9/3/2024

Pronunciation Tips

1

Pronunciation
is like
a sport...

2

Each sport

uses muscles

differently

3

4

5

6

9/3/2024

7

8

9

10

Each sport
uses muscles
differently

Pronunciation
is like a sport

Each sound in language
uses muscles
differently

11

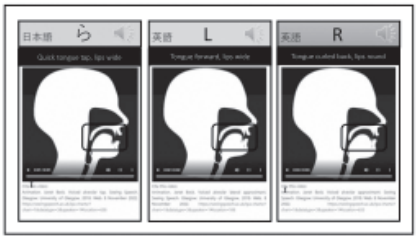
Each sound in language
uses muscles
differently

Pronunciation
is like a sport

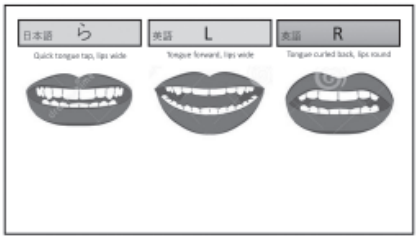
Pay attention: Lips
Tongue
Jaw

12

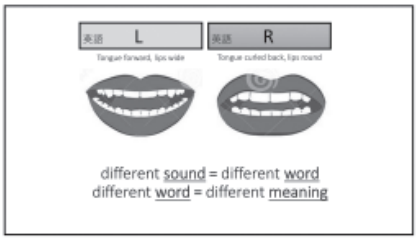
9/3/2024



13



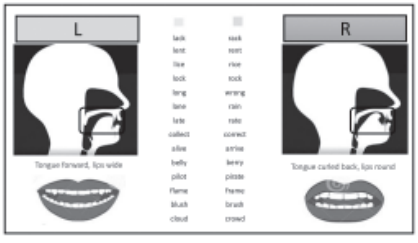
14



15

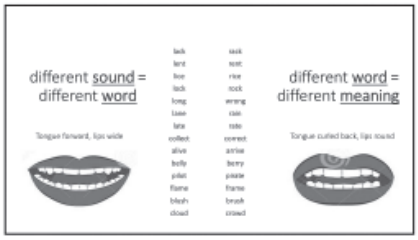
16

17



18

9/3/2024


different sound = different word
different word = different meaning

19