A Replication Study on English /l/ and /r/ Perceptual Training

Leah GILNER

要旨

本論文は，英語の /l/ 音と /r/ 音の弁別を対象とした，視聴覚的高変動音声知覚訓練（High-Variability Phonetic Training: HVPT）の再現研究の成果を報告するものである。Gilner（2025）は，約9週間にわたり総計30分程度という比較的短期間の介入によって有意な効果が得られたことを報告している。本研究の結果も，この種の訓練の有効性をさらに裏付けるものであった。先行研究と同様に，参加者は短時間の訓練後に有意な知覚的向上を示し，本介入の頑健性および集団や教室環境を越えた一般化可能性が確認された。さらに質的データの分析から，学習者は知覚的利益に加え，調音意識の向上，実用的な応用可能性，さらには動機づけの強化といった副次的効果にも価値を見出していることが明らかとなった。

キーワード：高変動音素訓練, 視聴覚知覚トレーニング, 知覚, EIL

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

1. Introduction

This paper reports findings from a replication study of an audiovisual high-variability phonetic training program (HVPT) targeting perceptual discrimination of

English /l/ and /r/[1] phonemes. A detailed description of the training program and findings from the original study can be found in Gilner (2025). As explained in the antecedent report, the instructional program was created with the aim of providing video-based materials for independent self-study. It is comprised of an introduction session and nine training sessions. The entire program takes about 30 minutes to complete, with each training session running just over two minutes. Results of the pre-test and post-test score comparison in the original study indicated statistically significant gains after the training. In addition, open-ended questionnaire responses revealed that the training helped participants notice a connection between how sounds are produced and how they are perceived as well as a high degree of variability across speakers.

The replication study reported herein was undertaken to see if comparable results would be obtained with a similar cohort of participants. An exact replication was carried out, meaning that the same methodology was followed as precisely as possible and the same steps were taken when analyzing the data. In this way, the present investigation goes some way toward assessing the reliability and generalizability of the original findings.

As previously mentioned, the training program implements an audiovisual (AV) approach to HVPT. Essentially, HVPT provides listeners with numerous samples of minimal pair words spoken by multiple speakers. The samples usually provide exposure to male and female voices of the same variety while instantiating various phonetic contexts. The visual cues complement the audio by illustrating the articulatory gestures underlying production of the target phonemes. Thomson (2018) conducted a meta-analysis of 32 HVPT studies and found that 97% reported significant improvement in learners' abilities to produce and/or identify the target sounds after training. Furthermore, it was observed that all of the studies provided evidence that training generalized to new words and new voices.

Moreover, previous research findings suggest that identification accuracy improves more when both audio and visual cues are available compared to when only one or the other is (Hardison, 2003, 2005; Shinohara, 2021; Shinohara & Iverson, 2021). Hardison (2005), for example, reports that AV input significantly improved identification speed and accuracy compared to audio only input. In the Hardison study, perceptual training was provided as part of fifteen 45-minute sessions including video recordings of American English speakers producing minimal pairs. L and R phonemes were among the target

sounds. While all participants in the study (university-aged learners from Japan and Korea attending an English program in the US) demonstrated more accurate and quicker identification after the training, it was observed that that Japanese L1 speakers made greater gains in identification speed and accuracy of words beginning with L and R phonemes than the other learners.

Shinohara (2021) provides corroborating evidence of an AV advantage in perceptual training. After ten 40-minute training sessions consisting of video-recordings of minimal pair words spoken by Standard Southern British English speakers, participants (child and adult Japanese L1 speakers) whose intervention included both audio and visual cues improved more than those in the visual only condition and those in the audio only condition. It is worth noting that participants were instructed to look at the speakers' mouth movements when the video clip was played, priming attention and increasing the likelihood of noticing the articulatory gestures associated with production.

In light of these encouraging results, the program under investigation here implemented a pluralistic approach to AV HVPT in an effort to supplement previous learning and experience. The intention was to capitalize on the benefits of this learning method while broadening the range of voices and accents that students are exposed to. This exposure is deemed beneficial given the rather narrow representation of the speech patterns provided by classroom learning and commercially available materials. This was made apparent by findings reported in Sugimoto and Uchida (2016). This study examined the required textbooks and accompanying audio recordings used in public junior high schools in Japan. Findings revealed that a North American accent was used to represent 46 of the 50 characters included in the collection of materials. This accent was used even for characters who were portrayed as being from India, Singapore, Japan, South Korea, China, Germany, and Cambodia. The accent used for three of the remaining four characters was identified as UK English by the researchers.

A survey of commercial textbooks commonly used at the university level reveals that so-called standard American and British accents are prevalent. Popular series such as Headway (Oxford University Press), Interchange (Cambridge University Press), and English Firsthand (Pearson) along with products sold by well-known Japanese publishers such as Gakken and Obunsha generally make similar choices when it comes to the accents represented in their audio material. This holds true for materials that specifically target pronunciation as well (Levis, 2016). A natural consequence of all this is that these

materials lack the variability encountered in every day interactions and hence students lack the opportunity to develop robust phonological categories and processing strategies.

With this in mind, a training program that could supplement previous classroom learning was developed. It is pluralistic in nature in that it includes nine different world English accents, namely, Australia, Hong Kong, India, Ireland, Nigeria, Philippines, Singapore, the United Kingdom, and the United States. Students hear male and female voices for three different varieties in each training session and voices rotate throughout the nine sessions. Complete details are provided in Gilner (2025) and pertinent facts will discussed in the Methodology section.

As was the case with the original study, the present investigation seeks to answer the following 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 invention along with weekly training test data will be used to answer RQ1 and RQ2. Open-ended questionnaire responses will be used to answer RQ3.

2. Methodology

Participants

Although the training was designed for independent study, it was completed by students from three intact courses under the author's supervision for the purposes of this study. Similar to the conditions in the original study, two of the courses focused on listening and speaking (Group AB: n = 19) and one on writing skills (Group W: n = 9). In line with the original study, the inclusion of Group W makes it possible to answer the second research question by comparing outcomes between groups receiving different types of classroom implementations. The most pertinent difference between the two

classroom implementations was that course-related activities for Group AB included extensive listening to conversations between speakers from different linguacultural backgrounds using English as a lingua franca. These conversations included individuals from Africa (Nigeria, South Africa), the Americas (Chile, Honduras, Mexico, US), Asia (China, Hong Kong, India, Indonesia, Mongolia, Sri Lanka), Europe (Belgium, Croatia, Portugal, Spain), Oceania (Australia), and the UK (England)[2].

Training materials

As previously mentioned, the 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, each one of which 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 high-quality recordings. Including the introductory session, the entire program takes approximately 26 minutes to complete.

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

The video materials, created with PowerPoint, have been designed to help students notice certain muscle movements associated with each phoneme. The Introduction session makes 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. Other visual cues include *Seeing speech* (Lawson et al., 2018) animated productions which illustrate the articulatory gestures corresponding to the target phonemes. Each phoneme is associated with a particular color scheme, key-word descriptors (e.g. tongue forward, lips wide), and a colored rectangle draws attention to the position of the articulators. In addition, a colored drawing of a frontal view of the mouth shape used for lip sync work (Mozart3737, n.d.) is associated with its corresponding phoneme. The Introduction closes with the presentation of 14 minimal pairs spoken in a US accent by a male and female voice and exemplifying word initial, intervocalic, and consonant cluster positions.

Each training session targets six minimal pairs, spoken by one male and one female voice in three different accents. Following each training session, a forced-identification test on the training items (and voices) was administered via Moodle. The results of these tests will be discussed in more detail below after the materials have been described.

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.

Training sessions 6 through 9 focused on visual cues associated with each phoneme. The training session videos included audio recordings of the target words as in Training 1 through Training 5. The only visual cue provided was the frontal view of the mouth

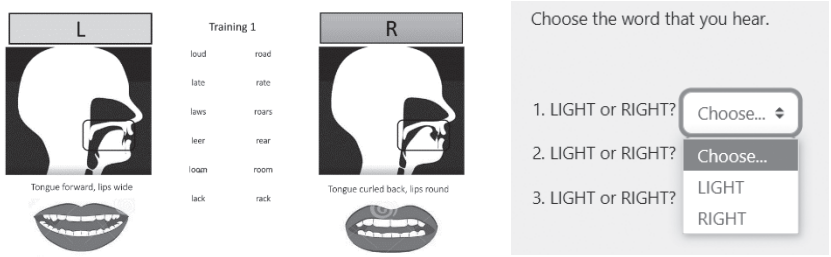

Figure 1. Training session (left) and testing presentation (right) format for sessions 1 through 5

shape 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.

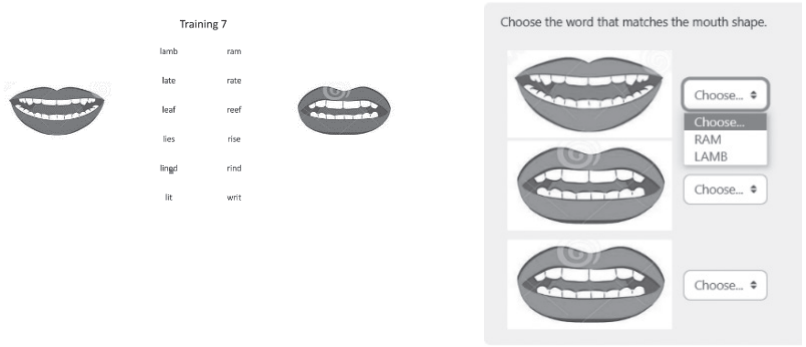

Figure 2. Training session and testing presentation format for sessions 6 through 9

3. Results

3.1 Pre-test and Post-test scores

Comparing average pre- and post-test scores indicates improvement in identification accuracy in both groups. Table 2 presents descriptive statistics based on average scores of the two tests. Group AB made an average gain of 5.10% and Group W of 6.13%. Standard deviation (SD) measures indicate that Group AB scores were relatively less variable on both pre-test and post-test compared to those of Group W, suggesting more consistent performance. In addition, the SD scores reveal a broader range on the post-test for both groups. The increase in SD from pre-test to post-test indicates the range of scores widens, possibly due to differing levels of improvement. Overall, the results suggest that both groups benefited from the training, with Group W showing slightly higher average gain despite variability. The trendlines in Figure 3 provide a visual display of these results.

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 AB	19	64.25% (25.70)	69.35% (27.74)	5.10%	8.32%	10.56%
Group W	9	68.33% (27.33)	74.45% (29.78)	6.13%	14.40%	21.79%

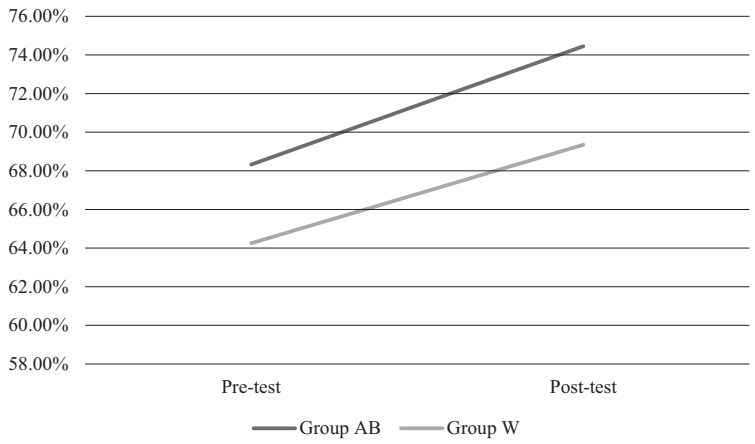

Figure 3. Pre- and Post-test trendlines for the two groups

A paired-samples t-test was conducted to compare scores from the two sets of data, Pre-test = Variable 1 and Post-test = Variable 2. The results showed a clear improvement in the average scores: Variable 1 averaged 26.71 (SD = 6.30), and Variable 2 averaged 28.89 (SD = 7.30), with a t-value of -2.80 and a p-value of 0.009 (two-tailed). The effect size (Cohen's d = 0.53) suggests a moderate practical difference, indicating that the change was not only statistically significant but also had a notable impact on the results. For our intents and purposes, these results suggest that the training had a measurable impact that would likely generalize to real-word contexts.

3.2 Training Test results

Table 4 presents the average training test scores for the two groups. These tests were taken immediately after viewing each of the training videos, once a week over the course of nine weeks. Scores of both groups increased over time and a relatively large jump in scores is observed between T5 and T6, coinciding with the change in task type. Group AB started T1 with an average of 60.39% and increased by similarly small increments up to T5. Average scores on T6 increased by more than 13% to 87.36%. Another significant jump of almost 10% is seen on the average score for T7 (97.92%). The average score on the last two training tests indicate mastery of the identification task. Table 3 illustrates the changes over the course of the nine sessions.

Table 3. Average test scores for the nine training sessions

	T1	T2	T3	T4	T5	T6	T7	T8	T9
Group AB	60.39%	61.22%	62.03%	64.08%	64.75%	87.36%	97.92%	99.53%	100.00%
Group W	65.28%	72.22%	74.22%	72.22%	64.17%	90.11%	93.89%	97.00%	96.00%

The average scores for Group W reveal a different trajectory. After the first training session, the average score was 65.28%. A noticeable increase of almost 7% is observed from T1 to T2, similar moderate increases of about 2% follow on the next two tests. Scores on T5 decreased by almost 8% to 64.17% and then rebounded by about 26% to 90.11% on T6. Scores increased by about 3% from T7 (93.89%) to T8 (97.00%). A decrease of 1% is observed from T8 to T9. Despite the variability, average scores from T6 onward indicate a high degree of identification accuracy. It is noteworthy that both groups obtained these high scores on the visual cue matching task tests. Figure 4 presents a visual display of the trajectories for both groups.

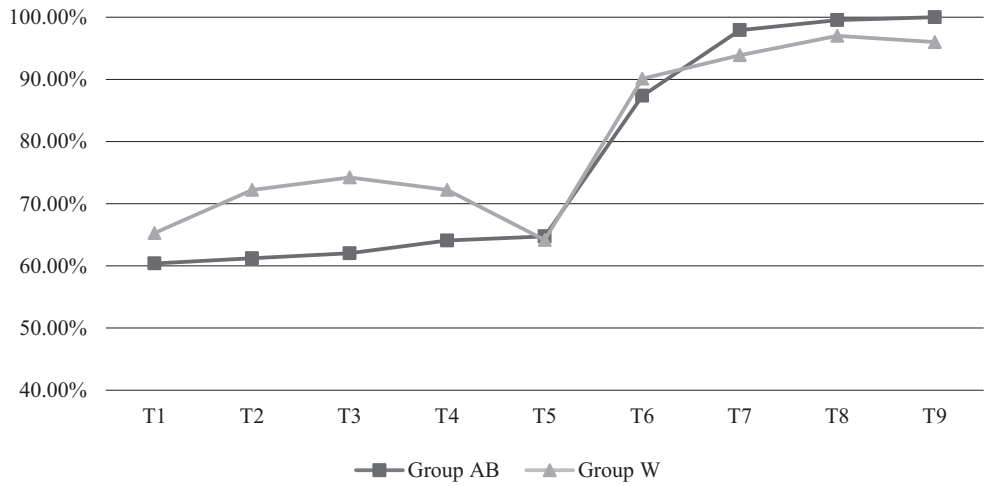

Figure 4. Test results for nine training sessions

3.3 Open-ended question responses

3.3.1 Thematic analysis

A total of 36 responses to an open-ended question were collected after completing the training program and taking the post-test. The precise wording of the question was as

follows:

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

Lexical and thematic analyses were carried out using KH Coder and Microsoft Excel. The responses were on average 58 words long, mentioning two learning points. A total of 76 learning points were identified and subsequently clustered into eight themes. It merits highlighting the fact that, while the training program focused on perception, themes encompass a wide-range of learning points including production techniques, affective factors, mother tongue influence, and speaker variability.

Table 4 displays the eight themes that emerged from the data, along with a representative sample response. The most commonly mentioned theme which occurred in 69.4% of the responses was *Listening and Pronunciation Improvement*. This theme captures general improvement in the perceived ability to better recognize and produce the target sounds, parse the speech stream, and catch key words.

Mouth and Tongue Movement was mentioned in 61.1% of the responses. This theme encompasses mentions made of articulatory gestures underlying production of the target sounds, including tongue position and movement as well as mouth shape.

Initial Difficulty and Progress was mentioned in 50.0% of responses. Respondents spoke directly of challenges in distinguishing the target phonemes particularly at the start of the training and observed gradual improvement over the course of the training.

Mentions of learning points related to *Practical Application* were found in 36.1% of the responses. This category includes learning points related to the practical application of skills beyond the training program; for example, specific mention was made of the TOEIC test and part-time job interactions.

Desire for Continued Practice, mentioned in 30.6% of the responses, reflects students' motivation or intention to further improve production and perception of the target phonemes.

Cultural/Linguistic Influence accounted for 25.0% of the learning points. In these cases, specific reference was made to the impact of one's mother tongue on the perception, production, and understanding of the target phonemes.

Learning points related to *Confidence and Motivation* were mentioned in 22.2% of the responses. These responses related how training produced an increased sense of

self-assurance in using the target phonemes as well as increased motivation to continue langauge study.

Variability in Speakers accounted for 13.9% of the learning points in the data set. This theme contains direct mentions of noticing how pronunciation of the target sounds changed across speakers. These observations were sometimes coupled with an acknowledgment of limited previous exposure to only "American English".

Table 4. Themes and Representative Responses from Open-Ended Questionnaire

Theme	Total Mentions	Percentage of Responses	Representative Example
Listening and Pronunciation Improvement	25	69.4%	"Through L and R training, I learned the importance of listening. At first, I could not tell the difference between similar sounds. However, after doing the training for a while, I started to notice small differences in pronunciation. It helped me understand English more deeply and improved my listening skills."
Mouth and Tongue Movement	22	61.1%	"I learned the shapes of the mouth to pronounce L and R. For me, it is still so hard to decide L and R from the sound. But I learned the shapes so maybe if I see the speaker's mouth it would be more clear to understand what they really want to say."
Initial Difficulty and Progress	18	50.0%	"In the beginning, I was not be able to differentiate L and R sounds. But, in the last training, I could answer all the questions correctly."
Practical Application	13	36.1%	"L&R training helped improve my listening and reading skills for tests like TOEIC. I learned to understand English conversations. This training helped me achieve a better score and use English in real life."

Desire for Continued Practice	11	30.6%	“This was my first time doing L and R training. I didn’t get a good score on the first test. However gradually I started getting good scores on the tests. I learned that pronunciation is different depending on where you live. It’s still very difficult for me to get all the answers right, but I want to keep practicing and try to get a perfect score.”
Cultural/ Linguistic Influence	9	25.0%	“Most of Japanese people have difficulty to recognize the difference between L and R sounds. So I think it is a good training to practice recognizing the difference between L and R sounds. I learned there are many pair of words that sounds same except for L and R sound.”
Confidence and Motivation	8	22.2%	“Through this training, I feel that not only my listening skills improved, but I also became more confident in pronouncing “l” and “r” clearly and correctly. Now I can notice the difference more easily and separate them when I speak. This made me speaking English more fun and existing.”
Variability in Speakers	5	13.9%	“I practiced L and R training with a lot of accents. The speaker was not just one but probably 5-6 people and each has their own voice and accent. From this, I realized listening to English spoken in various accents is important. Some of them were so easy to recognize L and R but other was so difficult. Japanese English education mainly uses American English so I get used to it. I will practice listening some other English accents this summer break.”

3.3.2 Lexical analysis

A hierarchical clustering analysis was undertaken with KH Coder to examine the responses in more depth. These analyses are based on co-occurrence patterns and reveal how terms form semantically and functionally coherent groupings. The bars on the left indicate relative frequency; the longer the bar, the more often the term occurred. The connecting lines on the right show how closely words cluster together and concurrently

how closely clusters are associated with each other.

Three main groupings emerged. The central grouping shown in Figure 5 displays three clusters. Cluster I-a contains the words which occurred most frequently in the responses, namely, *be, training, learn, listen, have, word*. This grouping highlights processes of acquistion and pedagogical training. Cluster I-b is adjacent to Cluter I-a suggesting a close conceptual relationship. Cluster I-b contains words associated with articulation and pronunciation including *difference, pronunciation, sound [noun], pronounce, mouth, shape, tongue*. These two groupings merge with Cluter I-c containing the words *understand, more, improve, skill, distinguish, help,* and *aware*. This cluster can be interpreted as relating to awareness and skill development. Taken together, this grouping encapsulates the learning points related to the most commonly occurring themes, **Listening and Pronunciation Improvement** and **Mouth and Tongue Movement.**

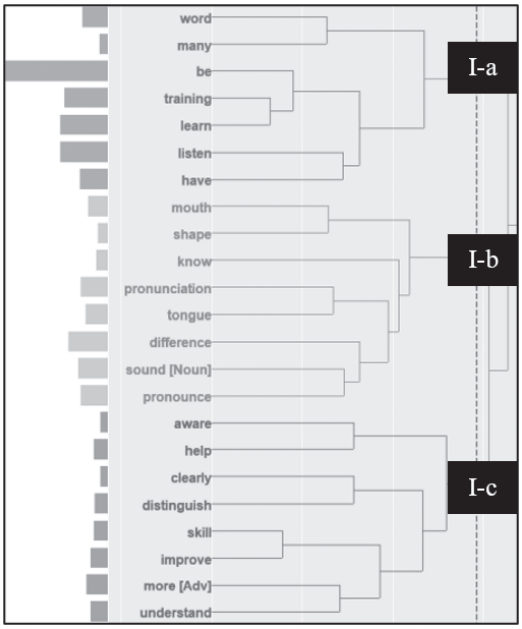

Figure 5. Grouping I in the hierarchical lexical analysis

Grouping I converges most immediately with Grouping II shown in Figure 6. Cluster II-b containing the words *help, clearly, distinguish, skill, improve, understand* is the anchor cluster. This cluster merges with the Cluster II-a (*time, score, test*) at an intermediate level, indicating that mentions of skill development and comprehension are

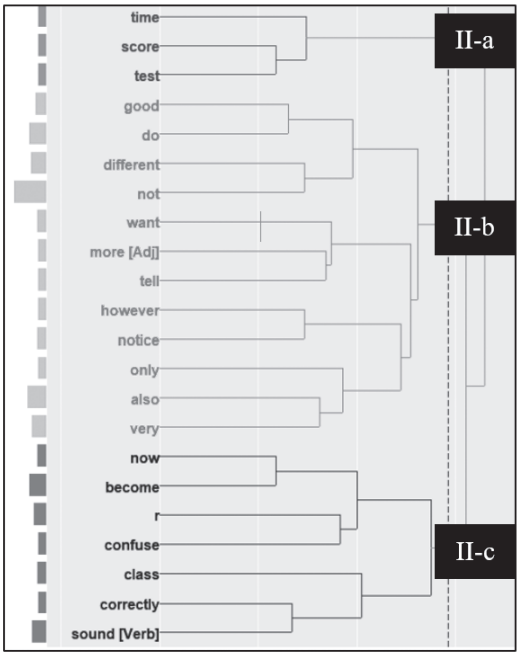

Figure 6. Grouping II in the hierarchical lexical analysis

closely tied to those of formal assessment contexts. This adjacency suggests that learners often perceive improvement in terms of measurable outcomes and view testing and scoring as integral to the process of monitoring skill gains.

Cluster II-b also merges with the Cluster II-c (*now, become, confuse, class, correctly, sound [Verb]*), linking awareness and improvement with classroom performance and error correction. This relationship suggests learners describe classroom training as a practice opportunity associated with becoming better able to discern correct and incorrect pronunciation. These clusters capture the learning points associated with the theme **Initial Difficulty and Progress**.

The third grouping is shown in Figure 7. The two clusters shown here converged relatively late with the anchor cluster I-a previously discussed. Cluster III-a containing *use, speak, English accent, important, lot, way,* lies adjacent to Cluster III-b containing *Japanese, speaker, recognize, practice, difficult, think, people.* This adjacency could be interpreted as reflecting an intersection between sociolinguistic awareness of English use and the challenges faced by L1 Japanese speakers. This suggests that responses often

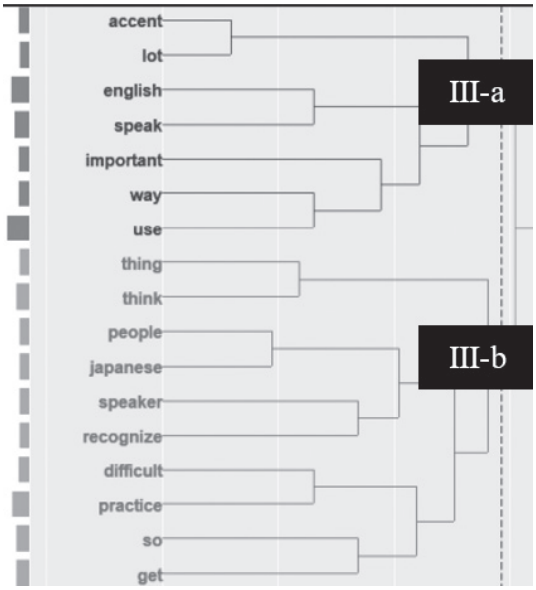

Figure 7. Cluster III in the hierarchical lexical analysis

connected the perceived importance of recognizing there are many English accents with the specific difficulties encountered by Japanese speakers. These two clusters encompass the themes **Cultural/Linguistic Influence** and **Variability in Speakers.**

Summing up, the central grouping anchored by the Cluster I-a, emphasizes training, learning, listening, and articulatory practice. This grouping converges next with the grouping anchored by Cluster II-b that encompasses awareness and skill development, assessment, and classroom performance. This relatively early convergence indicates a strong association between the processes of training and articulatory control with contexts of evaluation and classroom-based improvement. The later merger with the third grouping, centered on accent, English, and the difficulties faced by Japanese speakers, suggests that sociolinguistic concerns are recognized but remain more peripheral concerns.

4. Discussion

The present replication study set out to investigate the effectiveness of an audiovisual high-variability phonetic training program targeting the English /l/ and /r/ phonemes.

Three research questions guided the investigation: (1) whether students benefit from the independent training program, (2) whether results differ across groups with distinct classroom implementations, and (3) how students describe their understanding of English /l/ and /r/ after training.

With respect to RQ1, both quantitative and qualitative evidence indicates that participants benefited from the intervention. The pre- and post-test results demonstrated significant improvement in identification accuracy, with a moderate effect size suggesting practical as well as statistical significance. These findings align with previous HVPT research reporting gains in perceptual accuracy following short but intensive training sessions (Hardison, 2005; Shinohara, 2021; Thomson, 2018). The replication results therefore support the reliability of the original study (Gilner, 2025), reinforcing the conclusion that relatively brief, audiovisual self-study interventions can promote measurable improvements in the perception of L2 phonemic contrasts.

Regarding RQ2, differences between Group AB (listening/speaking courses) and Group W (writing course) were modest. While Group W showed slightly higher average post-test gains, the pattern of training test scores suggests that both groups followed similar developmental trajectories, with notable jumps coinciding with task-type changes. This convergence implies that classroom implementation, although distinct in focus, did not substantially alter the efficacy of the training program. The results suggest that the intervention is flexible and effective across different curricular contexts, a finding with practical implications for program design and adoption.

RQ3 was addressed through the analysis of open-ended responses, which revealed a broad range of learning outcomes beyond perceptual accuracy. The most frequently mentioned themes included improvements in listening and pronunciation, enhanced awareness of mouth and tongue movements, recognition of initial difficulty followed by progress, and increased motivation to continue practice. The prominence of articulatory and awareness-related themes underscores the role of visual input in heightening students' metacognitive understanding of sound production. Moreover, mentions of practical application (e.g., TOEIC, part-time work) and affective factors (confidence, motivation) suggest that students perceived the training as relevant and useful beyond the confines of the study. It is also noteworthy that speaker variability emerged as one of the themes in students' reflections, albeit less frequently than listening improvement or articulatory awareness. This finding is consistent with the pluralistic design of the

intervention, which deliberately incorporated multiple world English accents. While not the most prominent focus of the learning points, the recognition of variability suggests that exposure to diverse phonetic input heightened awareness of accent diversity and may contribute to the development of more flexible phonological categories.

The hierarchical clustering analysis further illuminates how students conceptualized their experience. Central clusters highlighted the interdependence of training, learning, listening, and articulatory practice, which merged closely with themes of awareness, skill development, and classroom improvement. By contrast, clusters relating to sociolinguistic awareness and L1-related challenges emerged as more peripheral. This structural pattern suggests that while students recognized the broader context of accent variability and L1 influence, they primarily described their learning in terms of practical, embodied, and classroom-relevant processes.

Taken together, these findings reinforce the benefit of multimodal, high-variability input in L2 phonetic training. They also highlight that, in addition to achieving perceptual gains, students developed articulatory awareness, and increased motivation. These affective and metacognitive dimensions are especially noteworthy since they can contribute to sustained engagement with pronunciation learning beyond short-term interventions.

5. Conclusion

This replication study provides further evidence for the efficacy of audiovisual HVPT targeting the English /l/ and /r/ phonemes. Consistent with the original study, participants demonstrated significant perceptual gains after brief training, confirming the robustness and generalizability of the intervention across groups and classroom contexts. Qualitative findings revealed that students valued the perceptual benefits of training as well as the heightened articulatory awareness, practical applicability, and motivational gains that accompanied it.

A distinctive feature of the present program was its pluralistic approach to HVPT, which exposed learners to a wide range of world English accents. This design choice contrasts with the relatively narrow input typical of classroom and textbook-based materials, and it appears to have broadened participants' awareness of speaker variability and accent diversity. Indeed, speaker variability emerged as a theme in students' open-

ended reflections, suggesting that pluralistic input was salient even if not as central as listening improvement or articulatory awareness. Such exposure may play an important role in preparing students for the phonetic diversity of real-world communicative contexts.

Limitations of the present study include the small sample size and reliance on intact classroom groups, which restricts the generalizability of the findings. Future research could expand the participant pool, include longitudinal follow-up testing to examine retention, and compare AV HVPT with other forms of pronunciation training. Additionally, further investigation of how learners transfer perceptual improvements to productive skills in spontaneous speech would be a valuable extension.

In conclusion, the study provides converging evidence that short, pluralistic audiovisual, high-variability training programs can contribute meaningfully to the acquisition of difficult L2 phonemic contrasts. Importantly, the findings emphasize that the benefits extend beyond perceptual accuracy to include increased articulatory awareness, practical readiness, and learner confidence, while simultaneously fostering awareness of speaker variability.

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

Gilner, L. (2025). A report on the design and outcomes of a perceptual training program targeting English /r/-/l/. *Aichi University Research Institute Working Papers in Language and Culture*, 79–100.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522. https://doi.org/10.1017/S0142716403000250

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), 3267–78.

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

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

Sugimoto, J., & Uchida, Y. (2016). A Variety of English Accents Used in Teaching Materials Targeting Japanese Learners. *International Symposium on Applied Phonetics (ISAPh 2016)*, 43–47. https://doi.org/10.21437/ISAPh.2016-9

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