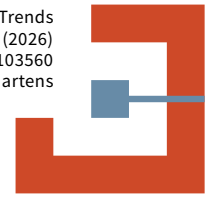
Trends

JALTCALL Trends
ISSN 2982-3749, 2(1), 103560 (2026)
https://doi.org/10.29140/jct.v2n1.103560



Castledown

Impact of multimodal phonetic training on the acquisition of American English vowels by native Japanese

Stephen Lambacher
Aoyama Gakuin University, Japan
lambacher@si.aoyama.ac.jp

William L. Martens
Macquarie University, Australia
bill.martens@mq.edu.au

This study investigated the effects of multimodal, computer-assisted vowel training on Japanese learners' recognition of American English vowels. Sixty first-year university students were assigned to either a multimodal training group using JSpace (an interactive vowel-space web tool) or a control group receiving auditory-only identification training. Both groups completed pretests and posttests measuring discrimination accuracy. Results showed significant within-group gains for both groups, although no statistically significant between-group differences were observed at posttest. The experimental group's JSpace selections demonstrated closer alignment with target formant (F1–F2) regions, particularly for /ɪ/, /æ/, and /ʌ/, whereas the auditory-only group showed stronger gains for high-front vowels and continued difficulty with low and back contrasts such as /ɔ/. These findings suggest that structured perceptual training – whether auditory only or multimodal – can support L2 vowel acquisition in computer-assisted language learning (CALL) contexts, with each approach showing distinct strengths.

Keywords: L2 vowel acquisition, multimodal training, phonological awareness, pronunciation pedagogy

Introduction

Second language (L2) speech learning can be challenging when the target language contains more vowel contrasts than the learner's first language (L1). For Japanese learners of American English (AE), the five-vowel Japanese system often leads to assimilation of AE vowels into the nearest L1 categories, hindering accurate

JALTCALL
Trends
Vol. 2 No. 1
2025 JALTCALL
Conference Papers
Special Issue

identification and production (Aoyama et al., 2004; Flege, 1995). This difficulty is especially noticeable for AE mid and low vowels such as /æ/, /ɑ/, /ʌ/, and /ɔ/, which lack clear Japanese equivalents (Lambacher & Martens, 2000). Nevertheless, adult learners retain sufficient neuroplasticity to benefit from structured perceptual training when provided with varied, high-quality input (Thomson, 2011).

Computer-assisted language learning (CALL) has become central to contemporary pronunciation pedagogy. Visual feedback tools, spectrographic displays, and automatic speech recognition systems can enhance segmental and suprasegmental accuracy (Hardison, 2004; Levis & Pickering, 2004; Neri et al., 2008). Such multimodal systems provide immediate feedback and promote learner autonomy. High-variability phonetic training (HVPT) further demonstrates that identification and discrimination practice improves perceptual accuracy (Lambacher et al., 2005; Martens & Lambacher, 2022). However, most L2 vowel training benefits primarily auditory recognition, without extending learners' phonetic awareness, which might impact their abilities to visualize and manipulate acoustic features in their vowel sound production.

Building on this work, the present study examines the effects of multimodal vowel-space training using JSpace, a web-based formant synthesizer integrating auditory and visual input. The goal is to determine whether multimodal training leads to greater perceptual gains than conventional auditory-only approaches within a CALL environment.

Methods

Participants

Sixty first-year Japanese undergraduates at Aoyama Gakuin University participated. Participants were randomly assigned to a multimodal-training group ($n = 30$) and an auditory-only control group ($n = 30$). Mean TOEIC scores averaged 610. None reported hearing or speech impairments, and all had prior English as a Foreign Language (EFL) instruction.

Stimuli

Training targeted seven AE vowels: /ɪ/, /ɛ/, /æ/, /ɔ/, /ʊ/, /u/, /ʌ/. Four vowels /i/, /o/, /e/, /ɑ/ were excluded due to their similarity to Japanese vowels. All vowels appeared in consonant–vowel–consonant (CVC) contexts. Japanese has a five-vowel system (/i, e, a, o, u/), which provided a context for presenting the seven targeted AE vowels that were contrasted with the learners' native L1 categories.

Procedure

Overview. Both groups completed pretests and posttests assessing AE vowel discrimination. The experimental group trained with JSpace, an interactive web-based tool allowing learners to explore the acoustic-phonetic vowel space by adjusting formant

frequencies (F1 and F2) to match target AE vowels. The control group completed auditory-only identification training. Each group trained for four weeks.

The JSpace system. JSpace is a browser-based vowel-space synthesizer derived from the VSpace platform, developed for phonetic research and instruction. It allows learners to manipulate first and second formant frequencies (F1 and F2) in real time. By clicking within a two-dimensional vowel chart, users generate synthetic vowels corresponding to specific acoustic coordinates. The interface displays the five Japanese vowel anchors (/i, e, a, o, u/) as reference points, enabling comparison between familiar L1 regions and AE targets. Immediate auditory feedback accompanies each cursor placement, supporting awareness of vowel height (F1) and front–back position (F2) (see Figure 1).

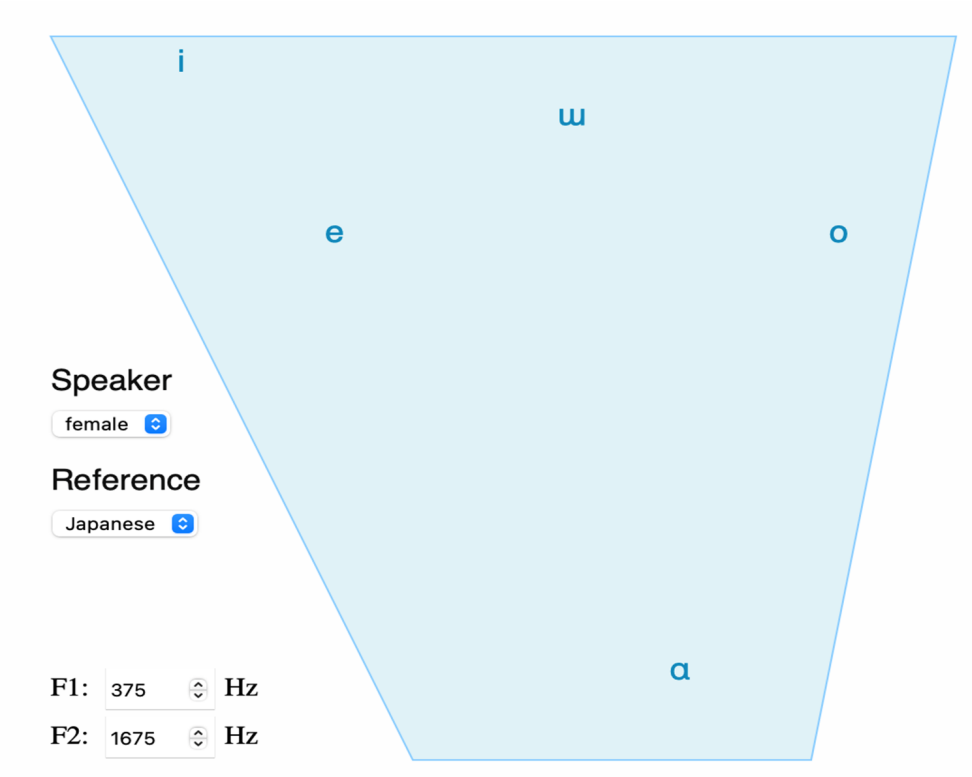


Figure 1. Web-based JSpace vowel-space interface used by experimental participants, showing five Japanese vowels as reference anchors for exploring nearby AE vowels.

Rather than simply replaying recorded tokens, learners actively search the continuous acoustic space to locate perceptual matches for target vowels. The selected F1/F2 values are displayed and logged, allowing systematic tracking of changes over time. By making acoustic distinctions visually and auditorily salient, JSpace supports learner control, multimodal input, and perceptual refinement while complementing traditional auditory training (see following link: https://mproctor.net/tools/jspace/).

Experimental group training. Participants received an orientation to JSpace. They

first produced vocalic glides between Japanese vowels to explore vowel-space transitions, linking acoustic and perceptual cues (see Figure 2).

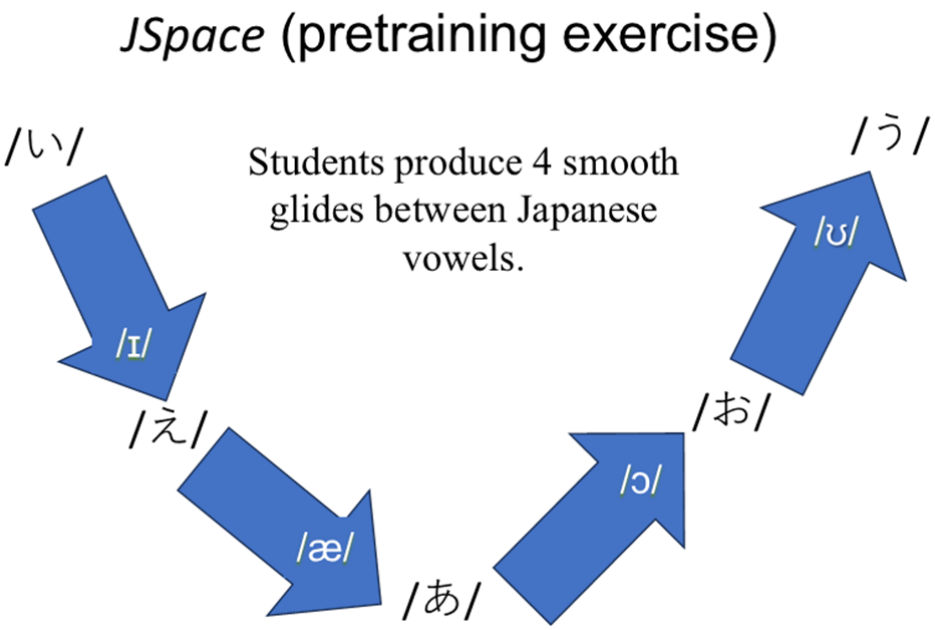
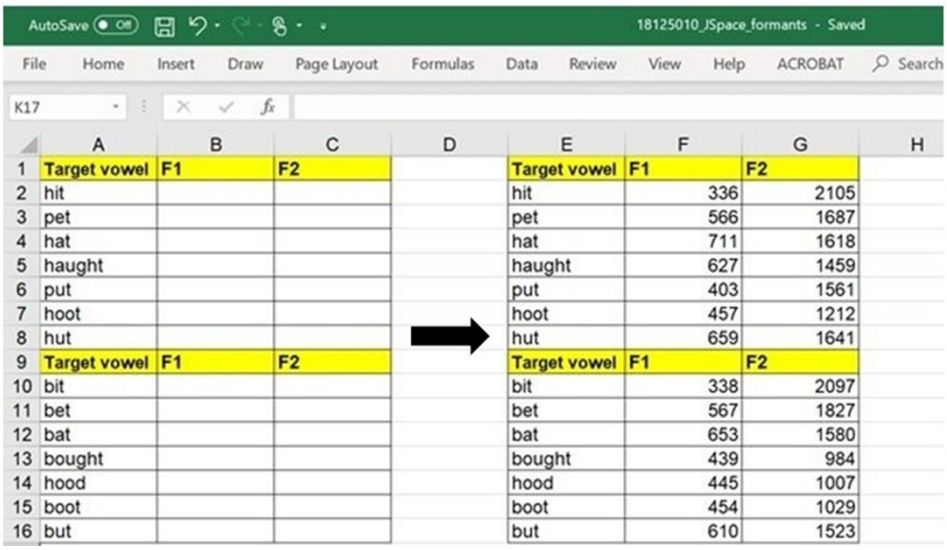


Figure 2. Orientation exercise illustrating how L2 learners explored intermediate vowels between Japanese vowels to familiarize themselves with the acoustic space.

Next, participants used the JSpace interface to manipulate vowel formant frequencies (F1 and F2) by clicking on the graphical user interface (GUI) to auditorily produce the target vowels. By adjusting the crosshairs and repeatedly clicking on points within the displayed vowel chart, they identified the location that generated the closest perceptual match to the target AE vowel. The JSpace tool displayed the corresponding F1/F2 coordinates for submission to analysis.

Students used Excel-based training logs that included native audio recordings of target CVC words. By listening to these samples, learners adjusted the position of their vowels on the JSpace display until the sound produced matched the model. This activity combined auditory, visual, and articulatory feedback; core principles of multimodal CALL-based pronunciation training (see Figure 3).

Training sheet (plotting F1/F2 values)



	A	B	C	D	E	F	G	H
1	Target vowel	F1	F2		Target vowel	F1	F2	
2	hit				hit	336	2105	
3	pet				pet	566	1687	
4	hat				hat	711	1618	
5	haught				haught	627	1459	
6	put				put	403	1561	
7	hoot				hoot	457	1212	
8	hut				hut	659	1641	
9	Target vowel	F1	F2		Target vowel	F1	F2	
10	bit				bit	338	2097	
11	bet				bet	567	1827	
12	bat				bat	653	1580	
13	bought				bought	439	984	
14	hood				hood	445	1007	
15	boot				boot	454	1029	
16	but				but	610	1523	

Figure 3. Excel-based recording sheet for logging participants' selected F1 and F2 values during training. Columns A–C show the initial blank cells to be filled in during the exercise, while Columns E–G show the completed spreadsheet with students' chosen formant frequencies copied from the JSpace interface after locating the target AE vowels.

Control group training. The control group completed auditory-only identification training once per week for four weeks. Twelve AE vowels were presented in CVC contexts (e.g., /bVd/, /bVp/, /hVd/), with 72 tokens per session (24 stimuli × 3 blocks). Participants listened to vowels produced by three native AE speakers and identified which vowel they heard by recording their responses in Moodle. Practice trials were provided to ensure that participants understood the task.

Testing tasks

Minimal-pair discrimination task. A pre/post minimal-pair discrimination task evaluated vowel recognition improvement. Nineteen CVC pairs were recorded by three AE speakers and embedded in the carrier phrase "Please choose the word ___." Examples included grin–green, sheep–ship, pan–pen, cot–caught, hit–heat, and hut–hot. Students completed practice trials before testing.

Results

Identification training (control group)

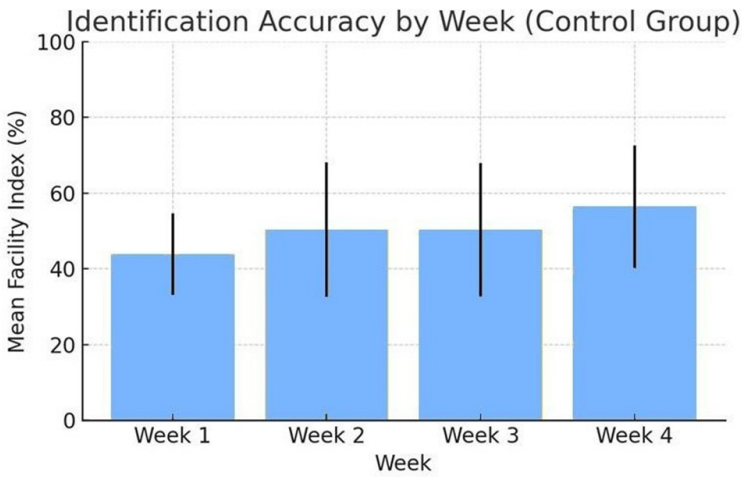


Figure 4. Mean identification accuracy (Facility Index %) across four weeks of auditory training for the control group. Error bars represent ±1 SD.

The control group's identification gains across four weeks of auditory-only training are shown in Figure 4. Identification accuracy increased over four weeks of auditory-only training. Mean accuracy rose from 43.9% to 56.4%, with significant improvement between Week 3 and Week 4 ($t(11) = 3.02$, $p = .012$, $d = 0.87$). Overall gain from Week 1 to Week 4 was large ($t(11) = 5.88$, $p < .001$, $d = 1.70$). Accuracy was highest for /ɪ/ (73%) and /u/ (65%), while /ɔ/ (21%), /ʌ/ (39%), and /ɑ/ (29%) remained the most difficult. The effect of vowel showed notable improvements for high and mid-front vowels (/i/, /e/, /ɛ/, /ɪ/), while back and low vowels (/ɔ/, /ʌ/, /ɑ/) remained challenging (see Figure 5).

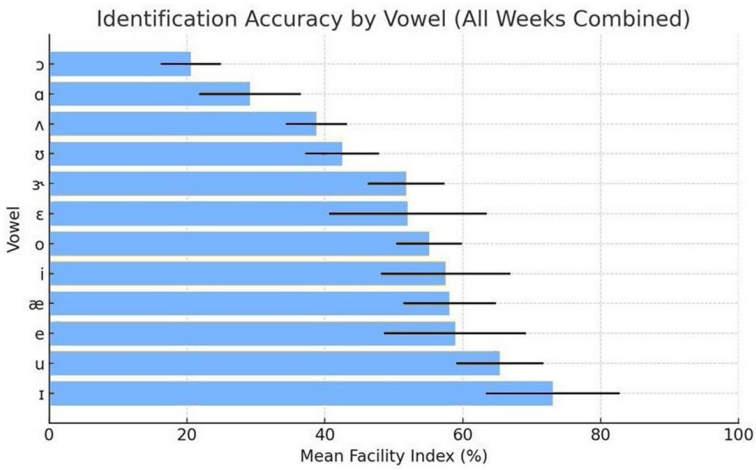


Figure 5. Mean identification accuracy (Facility Index %) by AE vowel (collapsed across all training weeks).

J-Space training (experimental group)

Formant frequencies the participants produced using the JSpace GUI are illustrated in Figure 6, which shows the acoustic outcomes of the multimodal vowel-space training. Mean formant values (F1–F2) for seven AE vowels showed convergence toward native-like target regions after four weeks of JSpace-based training. The overall vowel space expanded, particularly along height (F1) and front–back (F2) dimensions, with reduced dispersion across participants. Front vowels (/ɪ, ɛ, æ/) exhibited the largest positional shifts and the greatest reduction in variance, suggesting improved control of vowel height and advancement. Back vowels (/ʌ, ɔ, ʊ, u/) demonstrated smaller but consistent adjustments toward their target acoustic zones. These results indicate the integrated visual-auditory feedback enhanced vowel-space differentiation and precision, supporting the role of multimodal CALL tools in fostering perceptual-motor alignment in L2 speech learning.

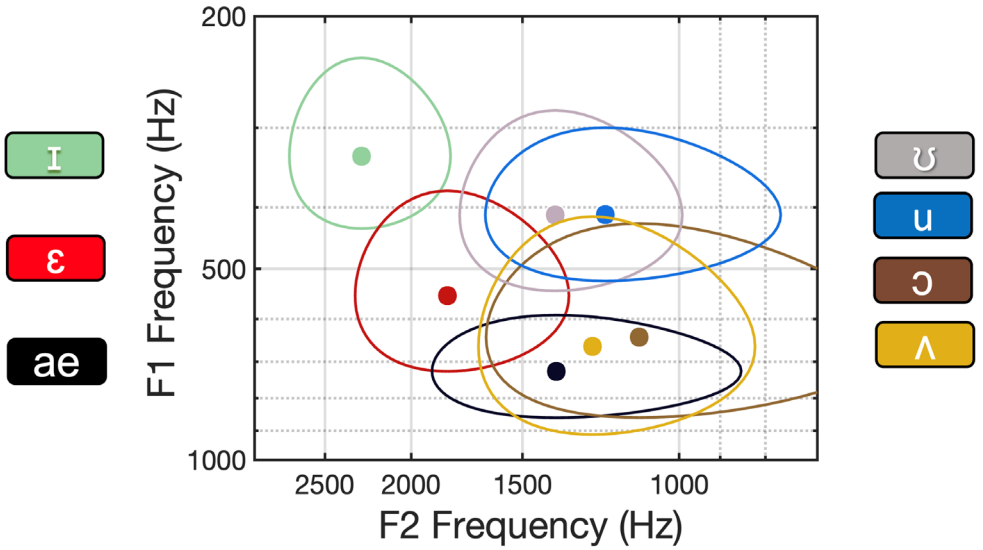


Figure 6. Mean frequencies for the first two formants of 7 vowels that produced a perceptual match to an AE target for a group of 30 students (solid circular plotting markers), and the associated 7 IPA symbols shown in the margins. The ellipses surrounding and centered on each mean graphically depict the standard deviations of F1 and F2 frequencies for each vowel.

Minimal-pair discrimination. Pre-post comparisons revealed modest but meaningful gains in vowel discrimination accuracy for both groups (see Figure 7). The control group improved from 86.13% to 88.57%, $t(31) = 2.79$, $p = .009$, $d = 0.49$, while the experimental group rose from 84.08% to 86.05%, $t(31) = 1.52$, $p = .139$, $d = 0.27$. No significant between-group difference occurred at posttest ($p = .476$). Accuracy was highest for high vowels (ship-sheep, hit-heat; 98–100%) and lowest for mid and low vowels such as hut–hot (44%) and ham–hem (64%).

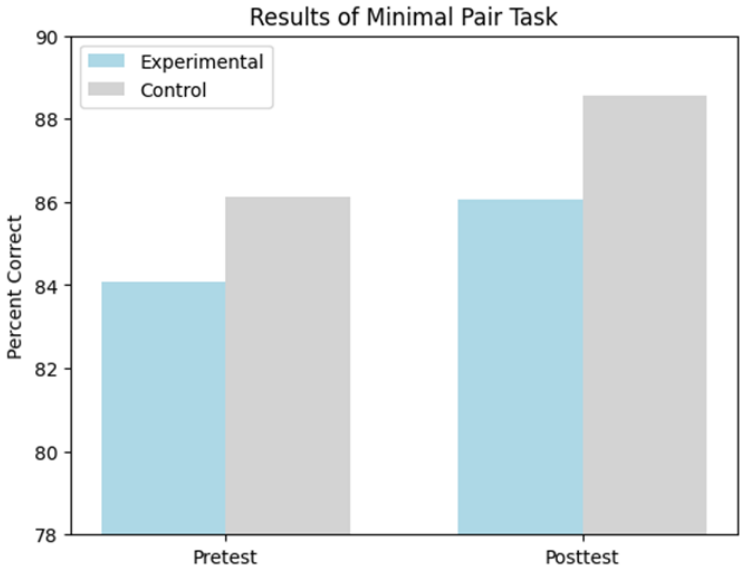


Figure 7. Minimal-pair discrimination task: Change in percent correct (from pretest to post-test) for two groups of participants, those in the "multimodal" experimental group versus those in the "audio-only" control group.

Conclusion

This study examined the consequences of multimodal vowel-space training using an interactive vowel-space tool (JSpace), in comparison with those of auditory-only identification training for Japanese learners' discrimination of AE vowels within a CALL framework. After four weeks, both groups improved in vowel discrimination, indicating that structured perceptual practice facilitates L2 vowel learning. Although no statistically significant between-group differences were observed at posttest, the JSpace group showed clearer vowel-space differentiation and reduced variability across several categories, suggesting greater awareness of acoustic-articulatory relationships. Overall, the findings support the pedagogical value of combining auditory input with visual formant exploration to enhance phonological awareness and learner autonomy, while highlighting that low and back vowel contrasts may require extended practice. Multimodal tools such as JSpace are therefore best viewed as a complement to auditory-only training, providing an additional pathway for self-monitoring and targeted perceptual refinement.

References

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 vowel learning. *Journal of Phonetics, 32*(2), 233–250. https://doi.org/10.1016/S0095-4470(03)00036-6

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology, 8*(1), 34–52.

Lambacher, S. G., Martens, W. L. (2000). A comparison of identification of American English vowels by native speakers of Japanese and English. In *Proceedings of Phonetic Society of Japan Annual Meeting.* (Chiba, Japan).

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26*(2), 227–247. https://doi.org/10.1017/S0142716405050140

Levis, J. M., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System, 32*(4), 505–524. https://doi.org/10.1016/j.system.2004.09.009

Martens, W. L. & Lambacher, S. G. (2022). Individualized identification training aids in the pronunciation of American English vowels for native speakers of Japanese. In *Proceedings of the 24th International Congress on Acoustics* (pp. 1–8).

Neri, A., Cucchiarini, C., & Strik, H. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental accuracy. *System, 36*(3), 423–435. https://doi.org/10.1016/j.system.2008.01.001

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

Author biodata

Stephen Lambacher is a Professor in the School of Social Informatics at Aoyama Gakuin University, where he has taught since 2008. He previously taught for 15 years at the University of Aizu. His research focuses on CALL and L2 speech acquisition.

William Martens is a Senior Research Scientist at the National Acoustic Laboratories and an Honorary Associate Professor at Macquarie University. He has held academic positions at the University of Aizu, McGill University, and University of Sydney, with research interests in psychoacoustics and speech perception.