

Munro, M. J., Derwing, T. M., & Saito, K. (2013). English L2 vowel acquisition over seven years. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, Aug. 2012. (pp. 112-119). Ames, IA: Iowa State University.

ENGLISH L2 VOWEL ACQUISITION OVER SEVEN YEARS

Murray J. Munro, Simon Fraser University

Tracey M. Derwing, University of Alberta

Kazuya Saito, Waseda University

Although cross-sectional research designs have been widely used in the evaluation of L2 phonetic learning, longitudinal studies of L2 speech production are rare. As a result, it is difficult to draw strong conclusions about the effects of language experience on L2 phonetic acquisition. This investigation of adult Slavic (Russian and Ukrainian) and Mandarin speakers tracks their English high vowel productions during seven years of residence in an English-speaking area. At the outset of the study, all participants had limited English oral proficiency. To evaluate phonetic learning, recordings of English vowels produced in controlled phonetic contexts were compared at the outset of the study, at one year, and at seven years. Vowel intelligibility was assessed through listener judgments in a blind identification task, and vowel accuracy was evaluated through acoustic measurements. While the results support the proposal that adults remain open to phonetic learning, they also indicate a dramatic slowing of the acquisition process by the end of the first year.

INTRODUCTION

Despite decades of research on adult second language (L2) phonetics, many aspects of the temporal development of L2 vowel and consonant production remain poorly understood. On the one hand, researchers have established that adults commonly do not learn to produce fully native-like L2 segments even after years of exposure to the L2. On the other hand, researchers have gained only limited insights into the amount of learning that actually does occur as a function of L2 experience and the time course of that learning (as opposed to instructed pronunciation learning). As part of an extensive longitudinal project examining oral language acquisition (see Derwing & Munro, 2013; Derwing, Munro, & Thomson, 2008), the present study aims at uncovering new details about the vowel acquisition process in adult ESL learners.

A much-needed type of work within the field of L2 phonetic learning is longitudinal research. Despite the preponderance of cross-sectional studies of segmental production comparing learners with different lengths of L2 residence (LOR), few solid conclusions about learning trajectories can be drawn, partly because LOR is a poor measure of L2 experience and also because of the problem of confounding factors in samples of learners drawn from different LOR populations. A further issue is that much of the work on segmental production has focused on phonetic “accuracy” – the extent to which L2 segments match those of the target L1 speakers. While that orientation can be useful in testing certain theoretical models, it tends to be misleading, and even counter-productive in applied phonetics, because phonetic accuracy is not a prerequisite for speech intelligibility (Derwing & Munro, 2009). For instance, inaccurate production of certain consonants such as /θ/, which participates mainly in low functional load phonemic distinctions, may have few serious communicative consequences for English

L2 speakers (Brown, 1991, Munro & Derwing 2006). Furthermore, if a segment such as /oʊ/ is produced intelligibly, but not entirely accurately, an improvement in accuracy may not yield any comprehension benefits for a speaker's interlocutors. In summary, a lack of acquisition of certain phonetic dimensions may not hamper L2 learners' communication skills, and improved accuracy on other dimensions may be communicatively irrelevant.

Although focused instruction is known to benefit L2 segmental production (e.g., Saito & Lyster, 2012), data on the time course of L2 phonetic development may be of use in identifying the types of learning that typically occur without intervention. Speech phenomena that fossilize early without instruction may be good candidates for classroom attention, provided they have a clear impact on intelligibility. At present, the results of studies of adult phonetic acquisition over time are mixed. Some research suggests that beyond a brief initial period of rapid phonetic learning, further L2 exposure has only limited effects on phonetic accuracy. Flege (1988), for instance, observed no cross-sectional difference in global foreign accent ratings between Taiwanese speakers of English with 1 year of US residence and those with 5 years of residence. Furthermore, Derwing and Munro (2013) found no longitudinal improvement in global accent ratings of Mandarin or Slavic speakers between their second and seventh year of Canadian residence. However, ratings of the Slavic speakers' comprehensibility (easy vs. difficult to understand) did improve significantly over the same interval. When specific segments have been considered, such as Japanese speakers' productions of English /ɪ/, learners with greater LOR have sometimes outperformed shorter-term residents (Flege, Takagi, & Mann, 1995) and sometimes not (Larson-Hall, 2006).

The current study aims to establish the learning trajectories for a particular set of segments – in this case high vowels – in English L2 learners after arrival in an English-speaking country. Here we address *intelligibility* through listener identifications and *accuracy* through acoustic measurements. This investigation extends earlier longitudinal work (Munro & Derwing, 2008) examining the vowel development of Mandarin and Slavic (Russian and Ukrainian) speakers during their first year in Canada. On arrival, all speakers in that study had low oral proficiency, and all were students in the same ESL program – one featuring no focused pronunciation instruction. Their productions of ten different vowels in CVC contexts were evaluated by both phonetically trained and untrained listeners. Significantly improved vowel intelligibility was observed during the first 6–8 months of Canadian residence, followed by a leveling off. This outcome appears to conform to Flege's (1988) proposal for global foreign accent. However, the same study revealed continued improvement over the entire year in both groups' performance on /ɪ/, which is missing from the phonemic inventories of both groups' L1s. Furthermore, /ɪ/ proved to be the least intelligible L2 vowel at both the beginning (5–31% correct) and the end (21–48%) of the study, with both groups of speakers tending to pronounce it as /ɛ/. Intriguingly, /ʊ/, which also does not occur in the speakers' L1s, showed somewhat better intelligibility than /ɪ/ at the outset, but no evidence of improvement over the year. A further finding was that vowels were produced more intelligibly in bVC than in pVC contexts, perhaps because of greater word frequency for the bVC items, which might lead to more exposure to native exemplars and, in turn, to more accurate perceptual representations on the part of the learners. Some questions that remained unanswered were whether further improvement would occur on /ɪ/ after the end of the first year and whether performance on /ʊ/ would remain unchanged. Another issue was whether the discrepancy in performance on bVC vs pVC would persist beyond the first year.

In view of the above outcomes from Munro and Derwing (2008), we focus our attention on the following research questions:

- Q1. In the absence of pronunciation instruction, to what degree will the adult learners of English improve in high vowel intelligibility and accuracy between years 1 and 7 of residence in their L2 environment?

Q2. If improvement occurs, which specific vowels will be affected?

Q3. Will the difference in intelligibility favoring bVC over pVC persist after 7 years of residence?

METHODS

Speakers

The speakers were 13 Mandarin and 18 Slavic-speaking adults who participated in Munro and Derwing (2008), all of whom were recent arrivals in Canada with low oral proficiency at the outset of the study. All had enrolled in the same ESL program, which featured no focused pronunciation instruction. Further details are given in Derwing, Munro & Thomson (2008) and Derwing and Munro (2013).

Speech Materials

Digital recordings were made of 10 bVC and 10 pVC productions from each speaker, where V = / i ɪ eɪ ε æ u ʊ oʊ ɑ ʌ /, and C = /t/ (except /k/ in the case of /bʊk/). At a number of testing points over the course of the study, tokens were elicited via a delayed repetition task in which the speakers heard the target words produced in the frame “The next word is ____,” and responded with “Now I say ____.” From the original recordings, the target words were excised from the sentence frame, normalized for peak amplitude, and saved as individual audio files for randomized intelligibility assessment and acoustic analysis. Although the speaking task was completed at multiple testing points, for the purposes of the present study, we will focus on productions from the one-year and seven-year points, and will compare them with productions from the outset of the study. Furthermore, we will be concerned only with the English high vowels / i ɪ u ʊ /.

Intelligibility Assessment

Vowel intelligibility was assessed by three phonetically-trained native English listeners, two of whom were the first two authors. The listeners heard the productions through headphones during a blind, randomized ID task in which they identified the closest native English vowel to the one heard. Items were presented via computer playback software (22.05 kHz, 16 bits), and identifications were made with screen buttons labeled with phonetic symbols. Buttons for all the possible vowel targets were available, along with ‘replay’ and ‘unknown’ buttons. Because of the large number of tokens, judgments were completed over several sessions.

Acoustic Assessment

For the high front and high back vowels, fundamental frequency (F0) and first and second formant (F1 and F2) measurements in Hz were made at the vowel midpoints from pitch tracks and formant tracks obtained through linear predictive coding in Praat (Boersma & Weenink, 2012). These were converted to Bark values, which correspond more closely than Hz measurements to human perception. Bark difference values (F1-F0, F2-F1) were then computed to reduce between-speaker differences, including gender effects, resulting from variability in vocal tract size.

RESULTS

Intelligibility

Identification data were converted to %-correct identification scores by tallying the number of times the high vowel tokens were labeled as the target vowel. Figure 1 provides mean identification scores pooled over /i ɪ u ʊ/ for the Mandarin and Slavic groups at two test times, with results broken down by initial consonant. These data were submitted to a mixed-design Analysis of Variance with first language (L1) as a between-groups factor and Time (T = 1 year, 7 years) and Initial Consonant (IC = b,

p) as within-group factors. Only the effect of IC proved statistically significant, $F(1, 29) = 4.685$, $p = .039$, $\eta_p^2 = .139$, indicating that high vowels in words beginning with /b/ were produced more intelligibly than those in words beginning with /p/. All other effects and interactions missed significance at $p > .1$

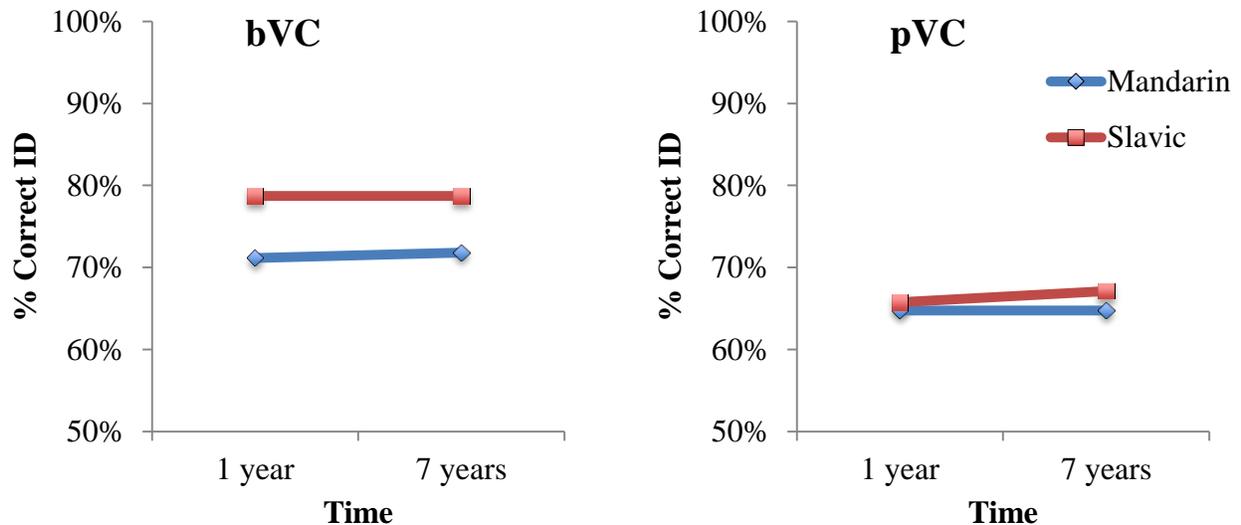


Figure 1: Mean intelligibility (% correct ID) of the two groups' high vowel productions at one year and seven years for bVC and pVC words.

Although the ANOVA provided no indication of an *overall* improvement in high vowel intelligibility, it is still possible that performance improved significantly on one or more vowels, but not on the others. Ideally, we would have liked to carry out statistical analyses for each vowel, but small cell sizes made such an approach inappropriate. Therefore, we present here an informal comparison across vowels in bVC context only, which is illustrated in Figure 2. To provide a fuller context, we include intelligibility data from the outset of the study (Munro and Derwing, 2008) with scores from the one-year and seven-year points.

For /ɪ/, improved intelligibility between the outset and the one-year point appears to have occurred in both groups, but there is no indication of meaningful improvement after one year. However, for /ʊ/, both speaker groups appear to show higher intelligibility at seven years. There is also slight improvement by the Slavic speakers on /i/, and by both groups on /u/.

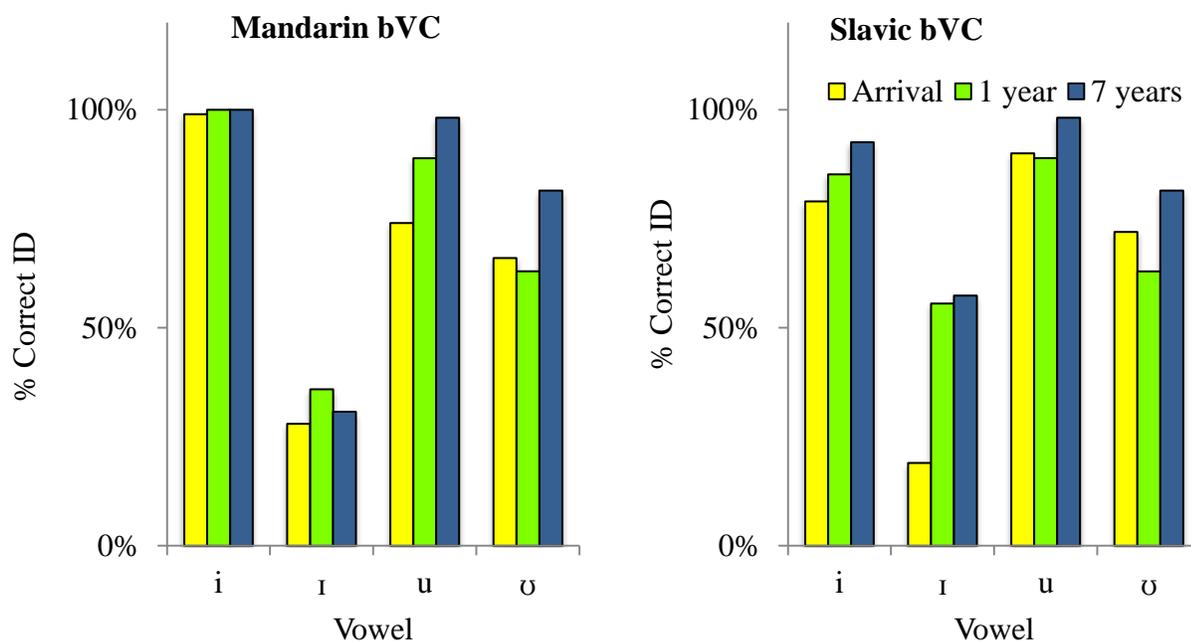


Figure 2: Mean intelligibility (%-correct identifications) of high vowels in bVC context produced by the Mandarin (left) and Slavic (right) groups at three times.

Acoustic Properties

To probe further the acquisition of the four vowels, acoustic data were informally evaluated. Mean values for the two groups are presented in Figure 3, with F2–F1 (Bark) on the x-axis and F1–F0 (Bark) on the y-axis. Bark scaling allows us to interpret the figure as an approximate representation of vowel height and advancement, with the arrows representing the direction and extent of change in tongue position for each vowel from the beginning of the study (in yellow) to the end of one year (in green) and until year 7 (in blue). Data were pooled from the bVC and pVC productions and are therefore not fully comparable with Figure 2. For both speaker groups, the clearest indication of change was for /ɪ/, which became higher and more advanced, particularly over the first year. Since that vowel was typically misproduced as /ɛ/, the direction of change is the expected pattern for improved intelligibility. The Slavic group, and to a much lesser degree the Mandarin group, showed some additional change in the production of /ɪ/ between the first and seventh years. Change in /ɪ/ was also considerably greater in the Slavic than the Mandarin group, a finding that fits well with the near-ceiling intelligibility on that vowel by the Mandarin speakers, and improved intelligibility over time by the Slavic speakers. For /u/, the Mandarin speakers showed higher and more forward articulations during the first year, followed by a regression back to the original vowel position, while the Slavic speakers showed higher and more forward productions during year one, followed by no further change. Finally, for /ʊ/, the Mandarin group showed lower, more back articulations over the first year, followed by a regression, while the Slavic group exhibited higher, more forward productions during year one, followed by a slight movement to higher positions.

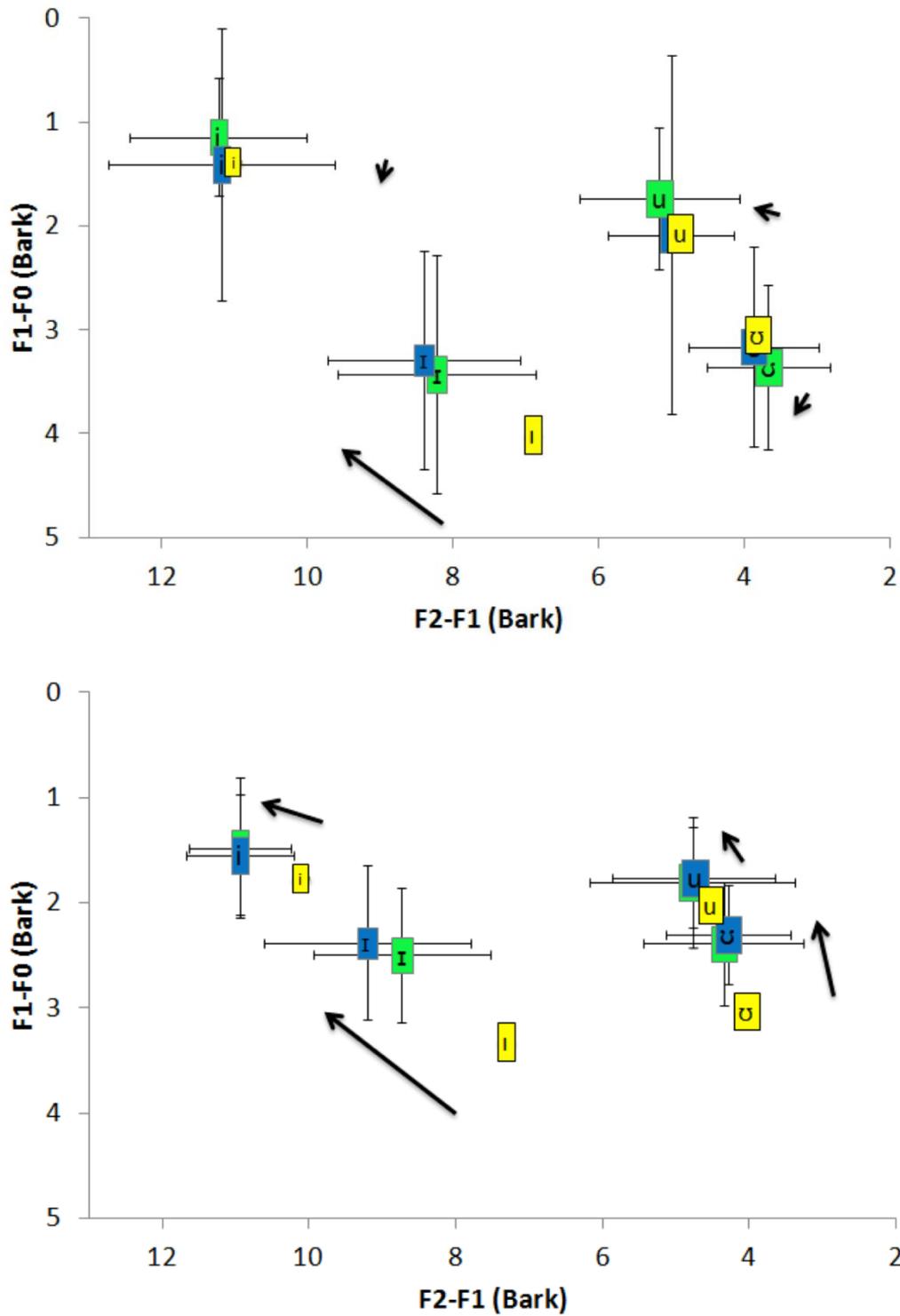


Figure 3: Changes in vowel formant frequencies for the Mandarin (top) and Slavic (bottom) groups from the outset of the study (yellow), to the 1-year point (green) and the 7-year point (blue).

DISCUSSION

In response to our original research questions, we can now make several observations.

Q1. Improvement in high vowel intelligibility and accuracy between years one and seven

Although there was no statistically significant change in overall high vowel intelligibility, an examination of the individual vowels suggests that some improvement in intelligibility did occur in both groups of learners. The acoustic data support this interpretation to some degree in that some productions appear to have improved in accuracy.

Q2. Specific vowels showing improvement between years one and seven

Listener evaluations suggest that the intelligibility of the Slavic speakers' /ɪ/ productions improved between years one and seven, while the acoustic data indicate improved accuracy (higher, more advanced productions) over the same time period. Data from the Mandarin speakers on /ɪ/ were not consistent with any improvement. While both groups appeared to show more intelligible productions of /u/ and /ʊ/ after seven years, the acoustic data suggested a regression in the Mandarin group toward original values. It is difficult to interpret these findings, partly because measurements of vowel formants taken at a single point do not reflect all the aspects of a vowel that determine its quality. (Vowel-inherent formant movement, for instance, is not captured.) A more detailed examination of the acoustic data in connection with specific vowel exemplars may lead to further understanding of the changes that the speakers actually implemented in their production strategies.

Q3. Intelligibility advantage for bVC vowels over pVC vowels

As was the case at the end of year one, vowels in the pVC context continued to be less intelligible than bVC vowels at the seven-year point. Whether this is a permanent aspect of L2 speech production should be explored through an examination of the productions of longer-term residents. A detailed study of a large set of vocabulary items varying in frequency should be conducted to determine whether there is a lexical effect for phonetic acquisition.

CONCLUSION

This longitudinal investigation of the acquisition of English vowels provides support for the view that the largest gains in segmental intelligibility and accuracy occur during the first year of residence in an L2-speaking area. However, the findings also suggest that phonetic learning does not cease altogether at the end of the first year. Rather, further improvements in intelligibility and refinements to production accuracy may occur naturalistically in some segments during the years that follow.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of Ron Thomson, Susan Morton, and the staff and students from NorQuest College. This research was funded by the Social Sciences and Humanities Research Council of Canada.

ABOUT THE AUTHORS

A former ESL teacher, Murray Munro is a Professor of Linguistics at Simon Fraser University. His applied phonetics research has appeared in a variety of international journals. Email: mjmunro@sfu.ca

Tracey Derwing is a Professor of TESL in the Department of Educational Psychology at the University of Alberta. She has conducted numerous studies of pronunciation and oral fluency development in L2 learners. Email: tracey.derwing@ualberta.ca

Kazuya Saito is an Assistant Professor of English in the School of Commerce at Waseda University (Tokyo, Japan). His research focuses on the role of experience and age in L2 speech learning in both naturalistic and instructed settings. Email: kazuya.saito@waseda.jp

Correspondence concerning this article should be addressed to Murray J. Munro, Department of Linguistics, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6, mjmunro@sfu.ca.

REFERENCES

- Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Retrieved July 2012 from <http://www.praat.org/>
- Brown, A. (1991). Functional load and the teaching of pronunciation. In Brown, A. (Ed.), *Teaching English Pronunciation: A Book of Readings* (pp. 221–224). Routledge, London.
- Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching* 42, 276–490.
- Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A seven-year study. *Language Learning*, 63, 163–185.
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics* 29, 359–380.
- Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America* 84, 70–79.
- Flege, J. E., Takagi, N., & Mann, V. A. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech* 38, 25–55.
- Larson-Hall, J. (2006). What does more time buy you? Another look at the effects of long-term residence on production accuracy of English /ɹ/ and /l/ by Japanese speakers. *Language and Speech* 49, 521–548.
- Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning* 58, 479–502.
- Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System* 34, 520–531.
- Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning* 62, 595–633.