An investigation of syllable position /l/ allophony in L2 English learners using Word Error Rate as an index of phonetic proficiency

Anisia Popescu[1,2], Lori Lamel[1,2], Ioana Vasilescu[1,2], Laurence Devillers[1,2]

[1]LISN, CNRS
[2]Université Paris Saclay
anisia.popescu@universite-paris-saclay.fr

Introduction. The present study compares the production of L2 English lateral consonant allophony of L1 French and L1 Japanese speakers using the word error rate (WER) of an automatic speech recognition system (ASR) output to determine phonetic proficiency. English dialects exhibit a syllable position allophony which opposes clear(er) [l] in onsets and dark(er) [ɫ] in coda and syllabic positions. The acoustic differences between the two English /l/ allophones are well documented: clear /l/ has high F2 and low F1; dark /l/ has low F2 and higher F1 (Sproat & Fujimura, 1993). A staple measure of /l/ darkness is defined as the difference between the second and first formants (F2 – F1). The difference is larger in clear /l/ and lower in dark /l/ (Recasens, 2012). Neither French, nor Japanese distinguish between the two lateral allophones: French is known to have clear [l] in all positions and the Japanese phonological system does not include a /l/ phoneme but has another phoneme corresponding to both the rhotic and the lateral consonant, generally considered to be a tap /ɾ/. Only a few studies have looked at the phonetic detail of the acquisition of the lateral allophony in L2 English (L1 Japanese: Nagamine, 2022; L1 French and/or Spanish: Colantoni et al. 2023; King & Feragne 2017, Barlow, 2014) and to our knowledge they focused only on/ l/-words read in isolation and in controlled carrier sentences without including proficiency in their analysis. The present study investigates the acoustic characteristics of /l/ production in read texts taking phonetic proficiency levels into account. Low proficiency L2 English learners are expected to show small or no differences in F2-F1 values in onset vs. coda/syllabic /l/. High proficiency L2 English learners should exhibit differences in F2-F1 values between the two varieties of /l/. This pattern should hold independent on the participant's L1. Nagamine's (2022) study on Japanese speakers found differences in production between the two allophones but included only participants with high proficiency. Colantoni and colleagues (2023) tested speakers of different French dialects that had been living in Canada and did not include proficiency level in their analysis. Self-reported proficiency levels based on foreign language assessment are usually a measure of global proficiency that does not necessarily correlate with phonetic proficiency. Having access to native speaker judgments is not always straightforward and can be time-consuming. We therefore opted for an alternative method, which, like native speaker judgments, relies on the acoustic characteristics of global accent rather than an overall proficiency level of second language acquisition – ASR for English with acoustic models trained on native speech production. The advantage relying on acoustic models trained on large amounts of data is that the acoustic qualities of different segments are based on a set of homogeneous features that reflect statistical patterns found in actual language use.

Methods. Participants were 22 French (12 female) and 18 Japanese (11 male) native speakers. All participants were recorded reading the same three beginner-, intermediate- and advanced-level texts (see Kobylyanskaya etal., 2023). The acoustic signals of the productions were forced aligned using WebMAUS (with English US as a reference language). All lateral consonant tokens were hand-corrected in Praat (Boersma & Weenik, 2022). A total of 40 different words containing singleton /l/s (17 in onset, 19 in coda and 4 in syllabic position) were included in the analysis. Segmental duration and formant measures (F1, F2 and F3) at the midpoint of the lateral were extracted for all tokens. The difference between the second and first formant (F2-F1) was used to assess darkness in the lateral.

Participants were split into three proficiency level groups (beginner, intermediate, proficient) using a classification measure based on word-error-rate (WER) calculated using the WER() function in Matlab from the output of an unbiased (no text transcription was available) ASR system (Lamel etal., 2011). The WER classification was benchmarked against the self-reported proficiency levels, which were based on language assessment tests (IELTS, TOEIC, TOEFFL, CAE etc.). /l/ darkness (F2-F1) was analyzed using linear mixed effects models (*lme4* package Bates etal., 2015). *Syllable position* (onset, coda, syllabic), *language* (French, Japanese), *proficiency* (beginner, intermediate, proficient), *sex* (female, male), flanking *vowel position* (front, mid, back) and interaction terms between *syllable position* and *language*, *proficiency* and *vowel position* were used as predictors. *Participant* and *word* were included as random factors with intercepts. A second model with *duration* of the /l/ segment as a response variable was run using the same predictor and random effects structure.

Results. /l/ darkness results show that Japanese speakers produce overall darker /l/s than French speakers (Est. - 215Hz p-value < .001). This could be a language specific difference or a by-product of the larger proportion of male speakers in the Japanese participant pool. For French speakers proficiency level plays a significant role in lateral allophone distinction: proficient speakers produce darker /l/ in coda (Est. -285Hz, p-value <.0001) and syllabic (Est. -298Hz, p-value <.05) position. Intermediate speakers produce darker /l/ only in coda position (Est. -173Hz, p-value <.01) but not in syllabic position. Japanese speakers present a different pattern, with speakers classified by the WER measure as

proficient not significantly differing from beginner speakers. Japanese intermediate speakers however show significantly darker /l/s than beginners in both coda (Est. -136Hz, p-value <.01) and syllabic (Est. -201Hz, p-value<0.05) positions. Both French and Japanese speakers exhibit coarticulation with the flanking vowel. /l/ produced in the context of a back vowel, independent of syllable position, is produced as darker when compared to /l/ produced in the context of front vowels (French: Est. -177Hz, p-value < .05; Japanese: Est- -163Hz, p-value <.05). Interaction effects between vowel and syllable position are not significant for either French or Japanese speakers. As to be expected *Sex* plays a significant role with male participants producing overall lower formants for both Japanese and French speakers.

Duration results show that /l/ duration plays a role only in the case of proficient Japanese speakers which produce longer /l/s in coda (Est. 17ms, p-value < .01) and syllabic position (Est. 30ms, p-value <.01). French speakers do not differentiate lateral allophones based on duration.
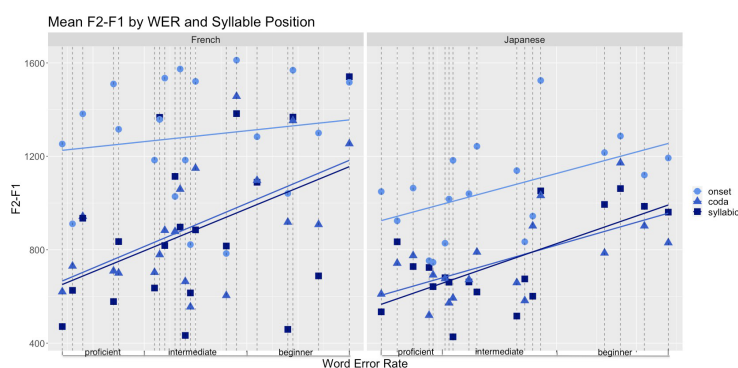

Figure 1: Mean F2-F1 measures per Word Error Rate values and Syllable position.
Fitted lines are provided for each syllable position level (onset, coda and syllabic /l/).

Discussion. The present study investigates the production of L2 English syllable position lateral consonant allophony in 40 speakers of L1 French and L1 Japanese. The acoustic analysis of the lateral allophones was related to the proficiency levels of the participants which was determined from the word error rate measured from the output of an automatic speech recognition system trained on native English data. Results show that while French speakers follow the predicted pattern of more proficient speakers producing more differentiated lateral allophones, the Japanese intermediate speakers produce the more differentiated allophones. Duration results also indicate a difference between French and Japanese speakers. In their study on /l/ allophony, Sproat and Fujimura (1993) found that /l/ is darker in longer rimes and suggested that the specific articulatory target of dark /l/ (i.e., tongue dorsum retraction) is more readily reached in longer rimes. This result was confirmed by Yuan and Liberman (2009) in a large corpus forced alignment study. The present study failed to find durational differences in French speakers but found durational effects for proficient Japanese speakers compared to beginner and intermediate speakers. This could be an indication that French and Japanese learners of L2 English tune into and adopt different phonetic strategies when producing phonetically different sounds in L2. The reported differences in the study do not appear when running the models using the self-reported proficiency levels, suggesting that using speech technology derived methodologies can provide more fine-grained phonetic classifications.

References

Barlow, J. (2014). Age of acquisition and allophony in Spanish-English bilinguals. *Frontiers in Psychology. 5. Article 288*

Bates, D., Maechler, M., Bolker B. & Walker, S. (2015) Fitting Linear Mixed-Effects Models Using lme3. *Journal of Statistical Software, 67(1), 1-48.*

Boersma, P. & Weenink, D. (2022). Praat: doing phonetics by computer from http://www.praat.org/.

Colantoni, L., Kochetov, A., & Steele, J. (2023). Articulatory Insights into the L2 Acquisition of English /l/ Allophony. *Language and Speech. 1-33.*

King, H. & Feragne, E. (2017). The effect of ultrasound and video feedback on the production and perception of English liquids by French learners. *Phonetics and Phonology in Europe. Cologne, Germany.*

Kobylyanskaya, S., Vasilescu, I., Augereau, O., Devillers, L. (2023) Vers la compréhension des difficultés de lecture en l2 à travers des paramètres acoustiques et de mouvement des yeux. 11ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain. EIAH2023

Lamel, L., etal. (2011). Speech recognition for machine translation in Quaero. In *Proceedings of the 8th International Workshop on Spoken Language Translation: Evaluation Campaign*, pages 121–128, San Francisco, California.

Moore, J., Shaw, J., Kawahara, S. & Arai, T. (2018). Articulation strategies for English liquids used by Japanese speakers. *Acoustical Science and Technology, 39*(2), 75-83.

Nagamine, T. (2022). Acquisition of allophonic variation in second language speech: An acoustic and articulatory study of English laterals by Japanese speakers. Proc. Interspeech 2022, 644-648.

Recasens, D. (2012). A cross-language acoustic study of initial and final allophones of /l/. *Speech Communication*, *54*(3), 368-383.

Sproat R. & Fujimura, O. (1993) Allophonic variation in Enlish /l/ and its implications for phonetic implementation. *Journal of Phonetics. 21(2), 291-311.*

Yuan, J. & Liberman, M. (2009). Investigation /l/ variation in English through Forced Alignment. Proc. Interspeech 2009, 2215