Have Four-Year-Olds Mastered Vowel Reduction in English? An Acoustic Analysis of Bilingual and Monolingual Child Storytelling

Paola Escudero[1], *Weicong Li*[1] and *Chloé Diskin-Holdaway*[2]

[1] MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia [2] School of Languages and Linguistics, The University of Melbourne, Australia

paola.escudero@westernsydney.edu.au, weicong.li@westernsydney.edu.au, chloe.diskinholdaway@unimelb.edu.au

Abstract

Recent studies show that late Spanish-English bilingual adults tend to under-reduce or over-reduce English vowels, with hardly any reduction in their Spanish vowels. We examined whether children acquiring Australian English and Spanish simultaneously show similar vowel reduction patterns. Bilingual four-year-olds' vowel acoustics (spectral and temporal) produced during story comprehension and retelling tasks were measured following previous methods used on child-directed speech. Unlike sequential bilingual adults, simultaneous bilingual four-year-olds mostly resembled monolingual peers in their English vowel reduction and acoustics. Findings are explained using the Second Language Linguistic Perception (L2LP) model, which uses input and language control to explain bilingual performance.

Index Terms: English vowel reduction, bilingual Spanish-English preschoolers, bilingual vowel production

1. Introduction

Vowel reduction is a well-known aspect of English phonology with a consensus on the fact that vowels in unstressed syllables are produced with less peripheral properties than stressed vowels [1-4]. What is less agreed upon is the specific properties of reduced vowels within and between varieties of English, with some researchers even referring to American English unstressed vowels as "targetless" [5] and highly dependent on position of the unstressed vowel in the word or sentence and the surrounding consonants [6]. Acoustic and articulatory studies of American English unstressed vowels have demonstrated that patterns of vowel reduction vary even within speakers [7].

There are fewer studies on vowel reduction in Australian English, but recent studies suggest that unstressed vowels are more likely to be reduced to the same central, mid vowel, in contrast to Southern British English where a higher and less central vowel is also used [8]. Adding to the variability in reduced vowels, [9] reports on formant values of Australian English vowels produced in spontaneous speech where many unstressed vowels are reduced to disappearance, e.g., "it's" produced as [ts] or [s] and the last syllable in "faces" and "places is produced as [s_z], with no vowel trace in any of these cases.

As mentioned by the authors in [9], English vowel reduction in natural connected speech may pose an even larger challenge not only for analysis but for an adult listener. The aim of the present study is to examine how this variability in vowel reduction affects language learning in bilingual populations, especially those for whom vowel reduction is not present in their language other than English.

Spanish-English bilinguals are a clear example of such bilinguals, with vowel reduction present in English but not in Spanish. The lack of vowel reduction in Spanish has been shown to contribute to high levels of foreign accentedness in the English speech of Spanish-English bilinguals [10]. Following [11] we define bilinguals as speakers of two languages who may have the same or different onsets of learning for the two languages. Bilinguals can be divided into simultaneous or sequential bilinguals, depending on the onset of language learning [11]. Within the sequential type, studies refer to early or late bilinguals, with the later commonly also referred to as second language (L2) speakers [e.g., 6].

Recent acoustic analyses have shown that late Spanish-English bilinguals tend to under-reduce or over-reduce some English vowels with respect to vowel formants and duration [12], while producing limited or no vowel reduction in their Spanish vowels [13]. High variability in the acoustic properties of English reduced vowels in both monolingual and bilingual adults has been highlighted in multiple studies, and explained by onset of learning, coarticulatory and lexical effects [6, 12]. For instance [14] found that early bilinguals differed less from monolinguals than late bilinguals, but that individual bilingual acoustic values varied more widely and displayed less group cohesiveness than monolingual values.

In this study, we focus on child rather than adult bilinguals. As mentioned in [15], studies comparing vowel acoustics in monolingual versus bilingual children are rare, with the authors mentioning a single study featuring Korean-English bilinguals [16]. Most studies on bilingual children's vowel production tend to use broad phonetic or phonological transcription [17, 18]. As mentioned above, vowel acoustic measurements are crucial to establish the extent of the variability of vowel reduction in both monolingual and bilingual children.

We could not find any previous acoustic study on English vowel reduction in children, but there are a few studies on bilingual children's vowel acoustics of English stressed vowels [see 19 for a review]. For instance, a recent study on Australian English measured the acoustics of /u/ produced by bilingual children with a variety of languages other than English (LOTEs) spoken at home. The authors found that bilingual children produced /u/ with a more retracted vowel than monolingual peers, with the bilingual pattern being correlated to the amount of use of the bilinguals' LOTE [20].

In the present study, we tested whether the development of vowel reduction in Spanish-English bilingual four-year-olds differs from that of age-matched monolingual peers, in line with previous studies with bilingual children and adults. In [15] it was reported that for American English stressed vowels, Spanish-English bilingual children behaved like late Spanish-English bilingual adults in that they did not exhibit the same

level of acoustic vowel contrast distinction as age-matched monolingual children. However, the authors showed that language dominance played an important role, with English dominant bilingual children resembling monolingual children more closely than Spanish dominant bilinguals.

We therefore examined both stressed and unstressed vowels to compare our results of simultaneous Spanish-English bilingual children living in Australia to previous results on children and adults living in the US. As for most adult studies, acoustic analyses of bilingual children's productions have used recordings of isolated words, but we follow [9] in using connected speech recorded within an elicited story telling task to represent a more naturalistic setting for vowel reduction.

2. Method

2.1. Corpus and participants

The recordings used in the present study are part of a large corpus of child language speech in bilingual and monolingual children [21], as part of a longitudinal project on non-societal language maintenance and enhancement in Australia [22]. The data presented here is part of the first wave of data collection for the Spanish-English cohort which aimed at measuring proficiency in their two languages. Proficiency was measured with a battery of psycholinguistic tasks, two of which targeted their ability to understand and retell the story conveyed in a colourful and engaging audio-visual eBook [23].

Data collection took place via Zoom with an experimenter conducting a session where children participated in the tasks with the help of a parent for session set up (for details on the online protocol see [24]. The story comprehension and retelling component of the session lasted for approximately 10 minutes, including the presentation of the 12-page eBook.

Here we report on vowel acoustic data from a subset of 8 children who participated in the first session of the longitudinal project. Table 1 shows the number of children in the monolingual and bilingual groups together with important demographics information, including children's age and language input. This information was gathered using a Qualtrics survey form sent to parents together with the electronic consent form and study information sheet. Data collection was approved by the human ethics committee of the first two authors' university (ethics approval number: H11022).

Table 1. *Participants' demographic information*

	Spanish-English Bilinguals	English Monolinguals
n	5 (4 females)	3 (3 females)
Mean Age (range)	4.45 (4.1-4.9)	4.55 (4.2-5.3)
Mean % exposure English (range)	40.0 (10-90)	100
Mean % exposure Spanish (range)	60.0 (10-90)	0

2.2. Data processing

Audio recordings sampled at 16kHz or above were first extracted from Zoom sessions run on desktop/laptop computers. The suitability and reliability of using audio recordings via Zoom for linguistic studies have been studied by [25, 26]. To gain orthographical transcription at utterance level, the audio recordings were processed using Whisper developed by OpenAI [27]. Whisper is an automatic speech recognition system trained on 680,000 hours of multilingual and multitask supervised data collected from the web. Resulting text files containing transcriptions and corresponding time information were converted to Praat [28] TextGrid files using MATLAB scripts.

Phonetically trained research assistants manually checked and corrected the transcriptions at utterance level. Using Whisper for ASR followed by manual review can substantially decrease human labour required for transcription, thus improving efficiency. Then the audio and transcription data were processed through WebMAUS [29] (Australian English model) to provide annotation at word and phoneme level. Boundaries and annotation were then manually checked by the research assistants who were instructed to use the automatic boundary unless it was obviously incorrect. We believe that the methodology of integrating the most recent technology such as Whisper in the workflow to speed up data processing could be widely applied to other linguistic research projects.

To annotate stressed and unstressed vowels, all the word tokens were first collated in a word list, which the research assistants went through and then manually marked the stressed/unstressed vowels. Then all the vowel tokens in the dataset were annotated as stressed/unstressed using the word list as a reference using a MATLAB script. Following [20], we annotated the Australian English vowel /ʉː/ followed by a lateral consonant as the allophone [ʊ], (e.g., SCHOOL), and the rest as [ʉː] (e.g., TWO).

2.3. Acoustic analysis

To examine vowel acoustics, duration and first and second format measurements were extracted in Praat. Formant measures were taken at 30 evenly distributed time points across the central portion of the vowel, starting from the 20% and ending with the 80% point [30-32]. For each vowel token, the series of 30-point vowel trajectories were smoothed using discrete cosine transform and averaged. Vowel tokens with fewer than 5 tokens were excluded. The final dataset included 1136 children's monophthong tokens: 718 by bilingual children and 418 by monolingual children. 714 were stressed syllables and 422 were unstressed syllables. Children's vowel acoustics were compared to those of the female adult speaker who read the eBook and conducted the Zoom session with the children, and who was a monolingual speaker of Australian English. She produced a total of 171 monophthong tokens.

As mentioned in the introduction, the main aim of the present study was to examine vowel reduction, specifically the amount of acoustic centralization in unstressed vowels, by both bilingual and monolingual children. To that end, we focused on the most peripheral vowels /ɪ, iː, æ, oː/, [ʉː], and [ʊ] and the most frequent unstressed vowels /ɪ, iː, æ, ə/, accounting for most child tokens and yielding approximately the same number of tokens for both vowel conditions (52% stressed = 371, 93% unstressed = 391). MATLAB and R were used for data visualization and statistical analyses on this subset of data, which we present below.

3. Results

Figure 1 shows ellipses for F1 and F2 values (mean and 2 standard deviations), together with individual tokens for stressed peripheral vowels and the most frequent unstressed vowel, namely /ə/. As expected of reduced vowels, /ə/ tokens have more centralised acoustic values than tokens of stressed

peripheral vowels in child and adult productions. Multivariate (F1, F2 and duration) pairwise comparisons yielded a significant difference between peripheral vowels and /ə/ (all *p*s < 0.01), except for monolingual stressed [ʉː] and [ʊ] and bilingual stressed [ʊ] and unstressed /æ/.

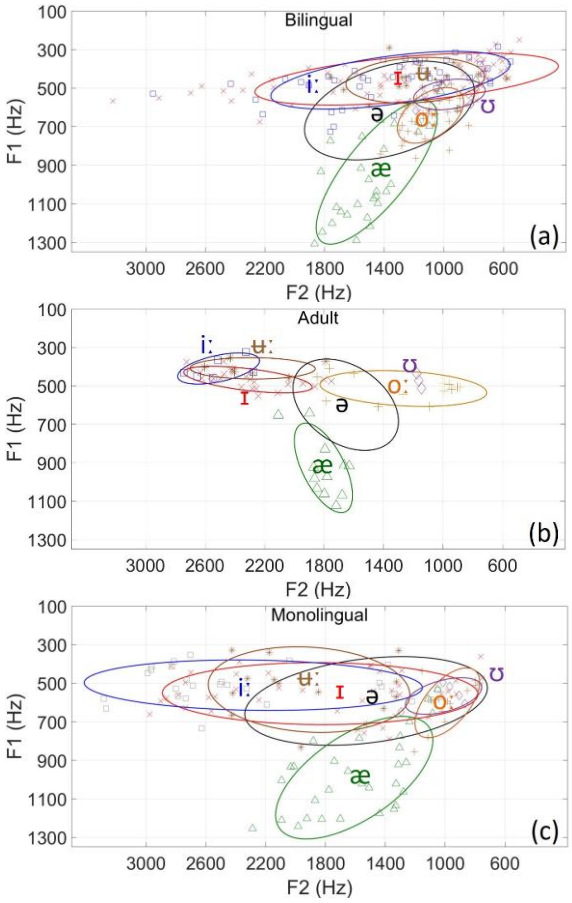

Figure 1: *Formant ellipses and data points for stressed peripheral vowels and /ə/. (a) bilingual children, (b) a monolingual adult, and (c) monolingual children.*

To further explore vowel reduction, we compared the acoustic values of the stressed and unstressed counterparts of /ɪ, iː, æ/. Separate linear mixed models for F1, F2 and duration with vowel, group, and stress condition as fixed effects and their interactions, and participants as random factors, revealed a significant interaction for group and vowel. Post-hoc pairwise comparisons showed that unstressed /æ/ had lower F1 values than its stressed counterpart for the adult (384 Hz lower) and bilingual children (268 Hz lower), bilingual unstressed /ɪ/ had lower F2 values (346 Hz lower), and monolingual unstressed /iː/ had lower F2 values (622 Hz lower) and longer duration (118 ms longer) (all *p*s < 0.05).

Children and adult productions seemed to differ only in some vowels and acoustic dimensions. Figures 2 and 3 show the largest differences in formant values with pairwise comparisons showing that the bilinguals and adult had lower F1 for unstressed /æ/ than monolinguals (at least 294 Hz lower), while bilinguals' F2 was lower than monolinguals and adult for stressed /ɪ, iː/ (at least 441 Hz lower) and [ʉː] (at least 546 Hz lower) and unstressed /ɪ/ (at least 727 Hz lower) (all *p*s < 0.01).

Regarding duration, adult vowels (both stressed and unstressed) were 58–243 ms shorter than those of children, while bilingual and monolingual children had comparable vowel duration, with only stressed /iː/ 108 ms shorter in monolinguals than bilinguals ($p = 0.03$).

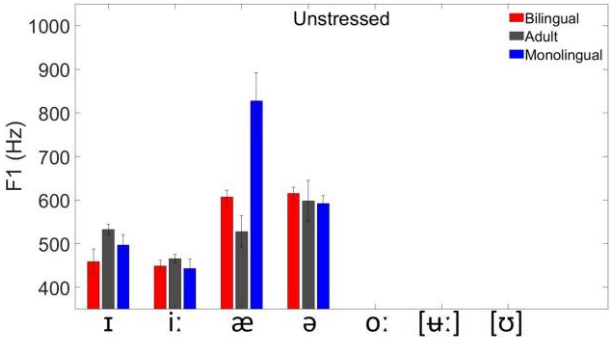

Figure 2: *F1 values (mean and standard error) of unstressed vowels.*

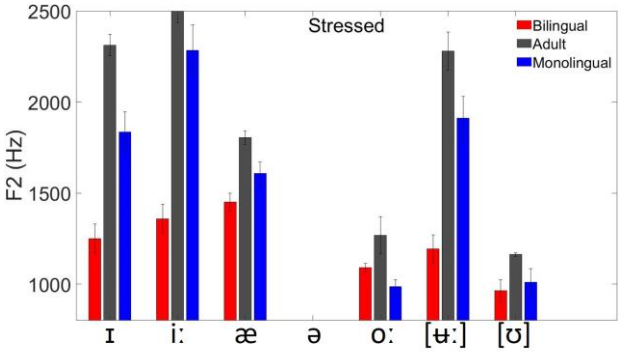

Figure 3: *F2 values (mean and standard error) of stressed vowels.*

4. Discussion and conclusion

As mentioned in the introduction, due to the lack of vowel reduction in Spanish, Spanish dominant bilingual children were expected to acquire English vowel reduction at a later stage in comparison to their English monolingual peers. This would be in line with previous studies with bilingual children and adults [15]. However, our findings demonstrate that 4-year-old children (even those with little exposure to English) have to a large extent mastered vowel reduction, as measured by the acoustic difference between their most frequent unstressed vowel /ə/ and their peripheral stressed vowels. Vowel reduction in children's productions was comparable to that of adult productions. Contrary to late bilinguals [12-14], bilingual children reduced unstressed /æ/ similarly to an adult, while monolingual children did not. This suggests that simultaneous bilingual children master vowel reduction more rapidly than monolingual peers.

The Second Language Linguistic Perception model (L2LP) [11, 33, 34, 35] proposes that input and the control of the different language modes are crucial variables for phonetic and phonological learning in bilinguals, predicting that amount of exposure and ability to inhibit the non-target language determine language learning [11]. In that respect, we also

observed some qualitative differences: compared to monolinguals who showed a strong split between the prelateral and non-prelateral allophones, bilinguals showed much less divergence in F2 values when producing [ʉː] and [ʊ]. This finding may indicate the interplay between the bilinguals' two languages, as suggested in [20] for bilingual children in Australia with diverse LOTEs.

Influenced by Spanish, high back /ʉː/ vowels produced by caregivers could lead to bilingual children's realisation of /ʉː/ in two ways: as a more back [ʉː] or possibly as a less retracted [ʊ], resulting in further divergence from monolingual peer productions. Additionally, four-year-olds may not be able to inhibit or control their dominant language, i.e., Spanish, and exhibit English vowel productions that are influenced by their Spanish vowel productions. Ongoing research aims at understanding the interrelation between individual language exposure and vowel production in the two languages of a larger cohort of bilingual Spanish-English children. In the analysis of this larger corpus, we will also consider the effects of position in the word, vowel length and prosodic context [36].

5. Acknowledgments

This study was supported by three grants from the Australian Research Council awarded to the first author (CE140100041, FT160100514) and to the first and last author (LP210300631). We thank all children who participated in the study and their parents who provided consent for their participation, Deeahn Sako for data collection, Minh Anh Tran, Sharmin Saba, and Kristy Bailey for help with manual segmentation and adjustment of the vowel data.

6. References

[1] N. Chomsky, M. Halle, The Sound Pattern of English. New York: Harper & Row, 1968.

[2] K. Crosswhite, "Vowel reduction", In: B. Hayes, R. Kirchner, and D. Steriade (eds), Phonetically-based Phonology. Cambridge: CUP 191–231, 2004.

[3] E. Flemming and S. Johnson, "Rosa's roses: reduced vowels in American English", JIPA 37, 83–96, 2007.

[4] L. Burzio, "Phonology and phonetics of English stress and vowel reduction", Language Sciences 29(2-3), 154–176, 2007.

[5] Y. Kondo, "Targetless schwa: is that how we get the impression of stress-timing in English?" Proc. Edinburgh Linguistics Dept. Conf. 63–76, 1994

[6] E. Byers, M. Yavas, "Vowel reduction in word-final position by early and late Spanish-English bilinguals", PloS one, 12(4), e0175226, 2017.

[7] M. Proctor, C. Y. Lo, and S. Narayanan, "Articulation of English vowels in running speech: a real-time MRI study." In The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences, International Phonetic Association, 2015.

[8] F. Cox and S. Palethorpe, "Rosa's roses – unstressed vowel merger in Australian English", Proc. 17th Australasian Int. Conf. on Speech Science and Technology (SST17). Australasian Speech Science and Technology Association (ASSTA), 89-92, 2018.

[9] G. Docherty, S. Gonzalez and P. Foulkes, "An acoustic study of the realisation of KIT in the conversational speech of young English speakers in Australia", Proceedings of the 20th International Congress of Phonetic Sciences, International Phonetic Association, 2023.

[10] J.E. Flege and O.S. Bohn, "An instrumental study of vowel reduction and stress placement in Spanish-accented English", Studies in Second Language Acquisition, vol.11, issue 1, pp. 35-62, 1989.

[11] P. Escudero P. and K. Yazawa, "The Second Language Linguistic Perception model (L2LP)," in Amengual, M. (Ed.). The Cambridge Handbook of Bilingual Phonetics and Phonology. Cambridge: CUP. Preprint. PsyArXiv, February 20, 2024. https://doi.org/10.31234/osf.io/qbx6z

[12] J.T. Conklin, O. Dmitrieva, Y.J. Jung and W. Zhai, "Acoustic characteristics of vowel reduction in advanced Spanish-English bilinguals", J. Acoust. Soc. Am. 148, no. 4_Supplement (2020): 2656-2657, 2020.

[13] J.T. Conklin, O. Dmitrieva, Y.J. Jung and W. Zhai, "Acoustic influence of L1 Spanish on L2 English vowel production", J. Acoust. Soc. Am. 150, no. 4_Supplement (2021): A42-A42, 2021.

[14] E. Byers, "Vowel reduction patterns of early Spanish-English bilinguals receiving continuous L1 and L2 input", Topics in Linguistics, 18(1), pp.17-31, 2017.

[15] S. van der Feest, G. Medina, E, Maryutina, I. Davidovich, T. Bloder, I. Barrière and V.L. Shafer, "Acoustic Correlates of Central Vowels in Russian-English and Spanish-English Bilingual Children", Proceedings of the 47th annual Boston University Conference on Language Development (BUCLD), 241-254. Somerville, MA: Cascadilla Press, 2023.

[16] S. Lee and G. Iverson, "Stop consonant productions of Korean–English bilingual children", Bilingualism: Language and Cognition, 15, 275–287, 2012.

[17] E. Jacewicz, R. Fox and A. Robert, "The effects of indexical and phonetic variation on vowel perception in typically developing 9- to 12-year-old children", Journal of Speech, Language, and Hearing Research, 57(2), 389-405, 2014.

[18] C.E. Gildersleeve-Neumann, E.S. Kester, B.L. Davis, E.D. Peña, "English speech sound development in preschool-aged children from bilingual English-Spanish environments. Lang Speech Hear Serv Sch. 39(3):314-28, 2008.

[19] J. Yang, "Vowel development in young Mandarin-English bilingual children." Phonetica. 22;78(3): 241-272, 2021.

[20] A. Gibson, F. Cox and J. Penny, "Acquiring allophony: GOOSE and SCHOOL vowels in the speech of Australian children", Proceedings of the 20th International Congress of Phonetic Sciences, International Phonetic Association, 2023.

[21] P. Escudero, G. Pino Escobar, M. Hernandez Gallego, Chloé, Diskin-Holdaway & J. Hajek, "The Little Multilingual Minds corpus: Educators' and children's speech in heritage languages", Workshop on community language corpora in Australia, Australian National University, Canberra, 2023.

[22] P. Escudero, G. Pino Escobar, C. Diskin-Holdaway, and J. Hajek. "Enhancing Heritage and Additional Language Learning in the Preschool Years: Longitudinal Implementation of the Little Multilingual Minds Program." OSF Preprints. September 16, 20204. doi:10.31219/osf.io/rvjcg.

[23] G. Pino Escobar and P. Escudero. "Vocabulary, Comprehension and Retelling in Multilingual Children: Age and Input Tell the Whole Story." OSF Preprints. September 16, 2024. doi:10.31219/osf.io/8e4nf.

[24] P. Escudero, G. Pino Escobar, C. G. Casey & K. Sommer, "Four-year-old's online versus face-to-face word learning via eBooks", *Frontiers in Psychology*, *12*, 450, 2021.

[25] C. Zhang, K. Jepson, G. Lohfink, & A. Arvaniti, A, "Speech data collection at a distance: Comparing the reliability of acoustic cues across homemade recordings", The Journal of the Acoustical Society of America, 148(4_Supplement), 2717-2717, 2020.

[26] C. Ge, Y. Xiong, & P. Mok, "How Reliable Are Phonetic Data Collected Remotely? Comparison of Recording Devices and Environments on Acoustic Measurements", Proceedings of Interspeech, pp. 3984-3988), 2021.

[27] A. Radford, J.W. Kim, T. Xu, G. Brockman, C. McLeavey and I. Sutskever. Robust speech recognition via large-scale weak supervision. In International Conference on Machine Learning (pp. 28492-28518). PMLR, July, 2013.

[28] P. Boersma, and D. Weenink, "Praat: doing phonetics by computer" [Computer program]. Version 6.4.13, retrieved 10 June 2024 from http://www.praat.org/, 2024.

[29] T. Kisler, U. Reichel, and F. Schiel, “Multilingual processing of speech via web services”, Computer Speech & Language, vol. 45, pp. 326–347, 2017.
[30] D. Williams and P. Escudero, “A cross-dialectal acoustic comparison of vowels in Northern and Southern British English”, J. Acoust. Soc. Am. 136, 2751–2761, 2014.
[31] H. Sarvasy, J. Elvin, W. Li and P. Escudero, “An acoustic phonetic description of Nungon vowels”, J. Acoust. Soc. Am. 147, 2891–2900, 2020.
[32] H. Sarvasy, W. Li, J. Elvin and P. Escudero, “Vowel acoustics of Nungon child-directed speech, adult dyadic conversation, and foreigner-directed monologues”, Front. Psychol. 13:805447, 2022.
[33] K. Yazawa, J. Whang, M. Kondo and P. Escudero, “Language-dependent cue weighting: An investigation of perception modes in L2 learning.” Second Language Research, 36(4), 557-581, 2020.
[34] X. Liu, P. Escudero, “How bidialectalism affects non-native speech acquisition: Evidence from Shanghai and Mandarin Chinese”. Applied Psycholinguistics 44 (6), 969-990, 2023.
[35] K. Yazawa, J. Whang, M. Kondo and P. Escudero, “Feature-driven new sound category formation: computational implementation with the L2LP model and beyond”, Front. Lang. Sci. 2:1303511, 2023.
[36] I. Yuen, F. Cox, and K. Demuth, “Three-year-olds' production of Australian English phonemic vowel length as a function of prosodic context.” The Journal of the Acoustical Society of America 135(3), 1469-1479, 2014.