Bilingual Preschoolers' Phonetic Variation Keeps Up with Monolingual Peers: The Case of Voiceless Plosives in Australian English

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

[1]School of Languages and Linguistics, The University of Melbourne, Australia
[2]MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia
chloe.diskinholdaway@unimelb.edu.au, weicong.li@westernsydney.edu.au, paola.escudero@westernsydney.edu.au

Abstract

We investigate phonetic variation in the English plosives /p t k/ of two groups of four-year-olds residing in Sydney, Australia: five Spanish-English bilinguals, and four Australian English monolinguals. Both groups are also compared with a monolingual adult. Nine categories of /p t k/ are identified across 901 tokens. No significant differences were found across groups, but /t/ exhibited the most variation, including variants such as a word-initial dentalized /t/, found only among the bilingual speakers. Results suggest that simultaneous bilingual children resemble monolinguals in their English plosives, with some minimal influence from their other language.

Index Terms: Australian English, Spanish, plosives, phonetic variation, simultaneous/early/late bilingualism.

1. Introduction

Phonetic variation in the voiceless plosives /p t k/ in Australian English (AusE) has been well-documented [1-4], with a particular interest in /t/, which has been shown to exhibit sociophonetic variation by dialect, region, age and gender [2]. There has been an emphasis on instances of glottalization, tapping (flapping), frication and deletion of these variables as compared to their canonical released variants, which are typically aspirated in word or syllable-initial position as [pʰ], [tʰ] and [kʰ]. However, these consonants display considerable variability, with e.g. [2] listing 10 categories of /t/ among AusE (mainstream and Aboriginal) adult speakers.

Despite interest in the speech of adults, research on the phonetic development of these plosives among children has been limited, with exceptions including some work on British English and American English with children as young as three years of age [5-7]. A review of studies in [8] finds that children around five years of age are beginning to produce gender-based variation in the realisation of glottalized stops in BrE. Exceptions to this gap in work on AusE children include [9], which studied gender-specific variation of plosives among primary school-aged children (ages 5-12) in Yarrawonga, Victoria. An earlier study on connected speech processes among Brisbane adolescents also documented some variation in /t/, particularly its weakening in particular phonological environments, such as intervocalically (prone to tapping/flapping) and in phrase-final position (prone to unreleased variants) [10].

In [9], six categories of /p t k/ were examined across 18 children at three year levels (Prep, ages 5-6, Grade Three, ages 8-9 and Grade Six, ages 11-12). For /p/, which had the least number of tokens overall, most variants for both boys and girls were canonical released tokens, followed by an unreleased variant. The fricated variant appeared at a higher rate in the speech of boys than it did for girls. For /t/, which had the most tokens due to it occurring statistically more often in English (see also [10: 37], both the fricated (including the fully fricated and affricated realisations) and the pre-aspirated variants appeared in girls' speech at a higher rate than that of the boys, which is consistent with findings of these variants in adult speech [see e.g. 11: 7]. Glottalization of /t/ (including both glottal and laryngealized realisations) also occurred at a higher rate for girls than boys overall. The tapped variant appeared at around the same rate for both boys and girls. For /k/, most were canonical released tokens, with gender-specific variation patterning similarly to /p/. There was more frication of /p/ among boys compared to girls. Both /p/ and /k/ had more pre-aspiration among girls as compared to boys.

Despite [9] including three age groups, with children as young as five, age was not a variable that was considered in the analysis, so little is known about the effect of age on the sociophonetic patterning of /p t k/ in the Australian context. Furthermore, there are only two previous studies that have examined phonetic variation (in vowels) among bilingual children in Australia: [12] and [13], who studied children living in diverse areas of Sydney, coming from a variety of Language Other than English (LOTE) backgrounds, but not Spanish. There have been no previous studies looking at variation in plosives among children with a LOTE background in Australia. Spanish is the tenth most spoken language after English in Australia, spoken by 171,378 people or 0.7% of the population, [14], so there is a sizeable population of children growing up bilingually with both Spanish and English in the country.

We focus on a population of four-year-old simultaneous bilinguals who have been exposed to Spanish and English at home and in the community. Previous developmental research on 3-4 year-old Spanish-English bilinguals in the U.S. context [15] shows that those with higher exposure to English have a lower error rate for English consonants overall than those with equal exposure to English and Spanish. However, the phonemes /p t k/, which occur in both languages, did not present any particular difficulty unless they occurred in consonant clusters or word-finally, where deletion was common [15].

Other work in the U.S. context on Spanish-English bilingual adults shows that only early bilinguals with age of onset 12 or younger produce plosives with voice onset time (VOT) contrasts resembling those of English monolinguals [16] (see also [17]), but late bilinguals have different VOT values. Our research questions are thus:

-What are the patterns of /p t k/ production among Spanish-English bilingual children compared to AusE monolingual children?
-How do these patterns compare to that of an AusE monolingual adult experimenter?
-How is /p t k/ variation influenced by utterance environment?

We predict that our bilingual group, all early simultaneous bilinguals, will have /p t k/ patterns resembling their monolingual peers, with some minimal differences expected for those with lower English exposure [15]. Since /p t k/ presents phonetic variability among somewhat older monolingual children [9], we investigate whether bilinguals also replicate this variability, or whether they have more canonical variants.

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 [18], as part of a longitudinal project on non-societal language maintenance and enhancement in Australia [19]. 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 a story in English conveyed in a colourful and engaging audio-visual eBook [20].

Data collection took place via Zoom with an experimenter (the 'adult' in our analyses) conducting a session where children participated in the aforementioned two tasks with the help of a parent for session set up (for details on the online protocol see [21]). The story comprehension and retelling component of the session lasted for approximately 10 minutes, including the presentation of the 12-page eBook. For four of the participants and the adult experimenter, we also include their speech as it occurred in the lead-up to the two tasks, and in a subsequent task measuring their ability to reproduce nonce words. This was to bolster the number of tokens for greater statistical power and comparability with [9]. Therefore, our data includes connected speech (short sentences) and isolated words produced by the children during these psycholinguistic tasks. The adult experimenter's data is mostly connected speech containing directives and also some isolated words e.g. *Ready?*

	Bilinguals	Monolinguals
n	5 (4 females)	4 (4 females)
Mean Age (range)	4.40 (4.2-4.8)	4.56 (4.2-5.3)
Mean % exposure English (range)	40.0 (10-90)	100
Mean % exposure Spanish (range)	60.0 (10-90)	0

Table 1. *Participants' demographic information*

Here we report on 901 tokens of /p t k/ extracted from a subset of nine children. Table 1 shows the number of monolinguals and bilinguals and their gender, age and language input, as reported by the parents in a Qualtrics survey. Data collection was approved by the human ethics committee of Author 2 and 3's university (ethics approval number: H11022). We conduct group-level comparisons (monolinguals vs. bilinguals) as well as discussion of individual behaviour, which we consider to be important with this small sample size.

2.2. Data processing and labelling

Audio WAV files were first extracted from Zoom video recordings. OpenAI Whisper [22] was used for automatic speech recognition (ASR). The use of Whisper for ASR can substantially speed up the transcription process and reduce the required human input. Text transcriptions at utterance level, together with time information, were then converted into Praat TextGrid files using MATLAB scripts. Research assistants (RAs) manually reviewed and improved the transcriptions for accuracy. Then forced alignment using WebMAUS [23] with the AusE model was applied on the audio recordings and transcription to obtain annotation at both word and phoneme levels. The /p t k / annotations at phoneme level were extracted as a separate tier in Praat and coded (see section 2.3). For each token, the position in the word (initial, medial, or final), and preceding and following segment (vowel, consonant, or pause) were also extracted based on the annotation at word and phoneme level, resulting in a total of 3^3, i.e. 27 utterance environments. Data visualisation and statistical analysis were performed using MATLAB and R [24].

2.3. Phonetic categories for /p t k/

Nine codes were used to categorize tokens and were influenced by [2] and [9]. Each token was subject to auditory and visual analysis, with corresponding spectrogram inspected in Praat. Two RAs conducted coding, which was manually checked by Author 1. Coding was primarily top-down, with RAs relying on the codes, but discussing categorisations with Author 1. Inaudible tokens, or where the speaker missed the target, e.g., pronouncing *school* as [stul], were excluded on the basis of an absent /k/ and thus not conforming to adult categories.

1. **Affricate**: a "closure followed by an /s/-like release (not aspirated), no burst-like characteristics" [t^s] [2: 62].
2. **Deleted**: no visible or audible presence of the consonant. "Difficult to distinguish phonetically from unreleased /t/" [10: 40], and as such, used for cases where the consonant would typically be released, but is deleted to due to a developmental/L2-like 'error', e.g. *ask* pronounced as [as].
3. **Fricated**: identified acoustically by "high frequency energy and lack of stop closure and release" [9: 66]. Not the same as /s/, better described as a "lowered /t/" [2: 62].
4. **Glottalized**: includes two realisations: glottal /t/ ("no formant transitions and the presence of creaky phonation on either side of the stop closure") and laryngealized /t/ ("lack of stop closure or release and presence of fully laryngealized voicing throughout the segment") [9: 66].
5. **H**: only applies to /t/ when it is pronounced like a [h].
6. **Other**: voiced /k/, dental fricative /t/, dentalized /t/, etc.
7. **Release**: the 'canonical' aspirated variant. Characterised acoustically and visually by a "period of full closure followed by burst. No voicing apparent" [2: 62].
8. **Tap**: identified acoustically by "a short closure phase and a short period of voicing" [9: 66]. Only occurs intervocalically and only for /t/.
9. **Unreleased:** no visible or audible presence of the consonant (see also 'deleted' category above).

3. Results

There were 203 /k/ tokens (23% of the dataset), 178 /p/ tokens (19%) and 520 /t/ tokens (58%) (Table 2), mirroring ratios in [9], for /k/ (25%), /p/ (13%) and /t/ (62%). In terms of numbers, the affricate, glottalized, release and tap categories appeared to differ between the three groups in the present study, but a Mann-Whitney U-test (MWU) found no significance.

group	variable	Aff.	Del.	Fric.	Glott.	H	Other	Rel.	Tap	Unrel.	Total
Biling.	k	1	7	4	1	0	5	85	0	2	105
	p	0	0	1	0	0	1	70	0	2	74
	t	34	11	1	59	2	22	98	12	3	242
Monoling.	k	0	9	5	1	0	2	55	0	1	73
	p	0	12	1	0	0	0	76	0	2	91
	t	23	10	4	30	1	5	90	21	15	199
Adult	k	0	0	1	0	0	1	23	0	0	25
	p	0	0	0	0	0	0	13	0	0	13
	t	6	0	0	15	0	1	48	7	2	79

Table 2. *Number of variants for /p t k/ for ten speakers, broken down by group (biling.=bilingual, monling.=monolingual, adult). Aff. = affricate; Del. = deleted; Fric.= fricated; Glott.=glottalized; Rel. = released; Unrel.=unreleased.*

3.1. Canonical release

Canonical release is the most prevalent category, particularly for /p/ and /k/. The adult had high rates for /k/ (92%) and /p/ (100%) and somewhat lower for /t/ (60.76%). The bilinguals had higher rates for /k/ and /p/, and the monolinguals for /t/ (Table 3). The children's rates resemble the adult (highest for /p/, lowest for /t/), but with lower overall release rates, indicating more phonetic variation as compared to the adult.

	Bilinguals	Monolinguals
/k/	82.9 (69.7–93.3)	73.4 (66.7–82.8)
/p/	88.6 (66.7–100)	85.3 (71.4–100)
/t/	39.6 (22.7–52.4)	47.5 (39.7–65.5)

Table 3. *Mean percentage (and range) of release by group*

A linear mixed effects model (LMM) compared release percentage (dependent variable) by two language groups (bilingual and monolingual) and consonants (/p, t, k/), modelled as fixed effects (independent variables), and participant as a random effect, and showed that only release for /t/ was significantly lower than for /k/ ($\beta = -43.3$, $p < .001$). No significant interaction between language group and consonant was found. Post-hoc pairwise comparisons for /k/ and /p/ showed bilinguals had (not significantly) higher release percentages than monolinguals, but the opposite for /t/.

3.2. Utterance environment

	/k/			/p/			/t/		
	i	**m**	**f**	**i**	**m**	**f**	**i**	**m**	**f**
pp	0	0	0	0	0	0	1	0	0
pc	5	0	0	7	0	0	11	0	0
pv	14	0	0	32	0	0	98	0	0
cp	0	0	6	0	0	0	0	0	30
cc	5	6	4	6	3	3	4	16	25
cv	8	23	0	15	4	1	38	19	5
vp	0	0	45	0	0	25	0	0	82
vc	3	44	8	8	1	10	16	35	56
vv	8	15	9	29	32	2	28	39	17

Table 4. *Tokens of /p t k/ by word environment (i: initial, m: medial, f: final) and preceding/following segment (c: consonant, v: vowel, p: pause)*

The majority of /k/ tokens were in word-final position, preceded by a vowel and followed by a pause, as in *truck*, and in word-medial position preceded by a vowel and followed by a consonant, as in *lunchbox* (Table 4). Both were frequent words, occurring 51 and 33 times respectively. Most /t/ was in word-initial position preceded by a pause and followed by a vowel, as in *two*, and in word-final position preceded by a vowel and followed by a pause, as in *but*. These patterns have implications for /p t k/ variation, discussed further below.

3.3. Variation in /p/ and /k/

/p/ had the least phonetic variation (Fig. 1), being mostly canonically released and with many word-initial tokens (Table 4). Both bilingual and monolingual children had some fricated and unreleased variants for /p/. Two monolinguals had deleted variants, but only in the nonce word task (section 2.1) in words like *sep* and *ballop* where it was difficult to ascertain whether the target was deleted or not met at all.

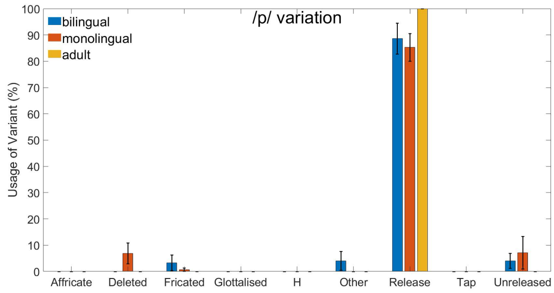

Figure 1: *Proportion of /p/ categories across three speaker groups (bilingual, monolingual, adult)*

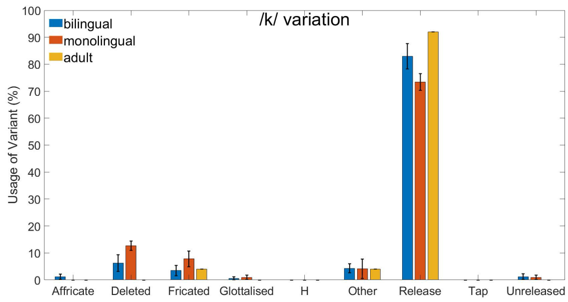

Figure 2: *Proportion of /k/ categories across three speaker groups (bilingual, monolingual, adult)*

/k/ had somewhat more variability than /p/ (Fig. 2), but no significant differences across groups and categories were found via MWU. Fricated /k/ was found across all groups, with some unreleased, deleted and glottalized tokens among the children. Monolinguals (12.33%) deleted somewhat more than bilinguals (6.67%), noting that many /k/ tokens were word-final and deleted in words such as *ask* and *desk*. All speakers had at least one /k/ in the ‘other’ category. All were voiced, except one palatalized /k/ for bilingual A018 in *take* and four ejectives for bilingual A017 in *okay*, *cake* (x2) and *truck*.

3.4. Variation in /t/

As well as being the most frequent of the three plosives, /t/ was also the most variable (Fig. 3), although no significant differences across groups and categories were found via MWU. The affricate category had somewhat higher proportions among the bilinguals (14.05%) and the monolinguals (11.56%) as compared to the adult (7.59%). In particular, the initial /t/ in *truck* was categorically affricated by the children, except one case of an ejective affricate by participant A017 and a case of deletion ([rʌk]), and epenthesis ([tʌra]) with a released /t/ by participant A016. The glottalized category was most prominent among the bilinguals (24.38%), followed by the adult (18.99%) and then the monolinguals (15.08%). However, the final /t/ context preceded by a vowel and followed by a pause had more glottalization by the monolinguals (53% in this context) than the bilinguals (39%) and was particularly prevalent in the phrase *don’t know*. Tapped variants were also more prominent among the monolinguals (10.55%) as compared to the adult (8.86%) and the bilinguals (4.96%). The ‘other’ category was prominent among the bilinguals, but this was mainly due to A034, who had 28.95% ‘other’, including a dentalized /t/ in words like *Tom* [t̪ɔm] and *to* [t̪u] and a voiced variant of /t/ in *light*.

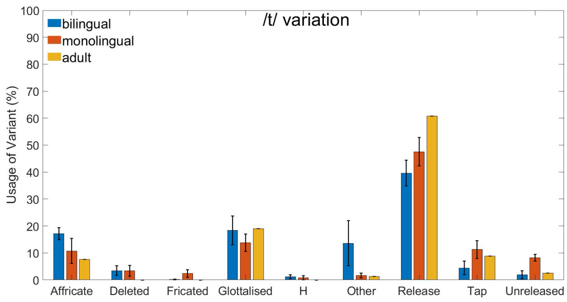

Figure 3: *Proportion of /t/ categories across three speaker groups (bilingual, monolingual, adult)*

4. Discussion and conclusion

We have presented phonetic variation in /p t k/ among a group of Spanish-English bilingual children as compared to AusE-speaking children and one adult. Overall, we find little quantitative evidence for major differences between the groups. Apart from some minor differences, the children appear to have mastered adult-like phonetic variation in plosives, and bilingual children match monolingual peers. We note that the adult had less variable /p t k/ overall as compared to the children and to other studies on AusE-speaking adults [e.g., 2]. As experimenter, they were likely engaging in careful speech to come across as clear and articulate. Child-directed speech has also been found to enhance discriminability of stop consonants via exaggeration of VOT contrasts [25].

/t/ presented the most phonetic variability across all groups, consistent with other work on /t/ in AusE [e.g. 2]. However, the ‘other’ category for /t/ was notable among participants such as A034, who also had one of the highest rates of Spanish listening (90% of the time) and Spanish speaking (95%; aggregate exposure scores in Table 1). These variants, particularly a dentalized /t/ in syllable-onset position, can be traced to direct influence from the Spanish denti-alveolar consonant. Qualitative observations among other participants with relatively high Spanish listening/speaking (A019 – 50%; 50%; A016 – 86%; 94%) suggest other influences of Spanish, such as coda /s/ deletion and epenthesis in consonant clusters (also found in [15] in the U.S. context).

We note some further developmental observations: the word *lunchbox* presented some difficulty, with several bilingual children, e.g. A019 and A016, variably deleting the final /s/, pronouncing the word as [lʌnʃbɔk]. Participant A016 had five final /s/ deletions across their tokens of /p t k/. Participants A016 and A034 also introduced vowel epenthesis in the consonant cluster ([lʌnʃəbɔks]) in some *lunchbox* tokens, which could also be an influence of the word boundary (*lunch* and *box*). While epenthesis is common in children’s speech, it is worth noting that this phenomenon only occurred among the bilingual children in this sample. However, the monolingual children also exhibited some likely developmental differences in their deletion of the final /p/ in nonce words like *sep* and *ballop*, at similar frequencies to the ‘other’ category for /t/ among the bilinguals.

In comparison with plosives in AusE monolingual children [9], we find lower rates of unreleased variants and somewhat higher rates of released variants, particularly for /p/ and /t/. However, this could be explained by the fact that [9] included more connected speech (elicited through “spontaneous conversation and interactive games and activities”) than our dataset, whereas we had children speaking both in short sentences and using standalone words, usually in response to a question as part of a task led by the experimenter. Despite this, [9] had similar proportions of token numbers of /p/:/t/:/k/ as in our study. Other differences include higher rates of /t/ glottalization in our sample: as high as 24% for the bilinguals versus a high of 10% in [9]. Conversely, our participants have lower rates of /t/ and /k/ frication than in [9] and compared to adult AusE speakers [1, 26]. Our rates of tapping are also lower than in [9] (rates close to 20%), but our monolinguals and adult (8-10%) have rates closer than the bilinguals (4.96%).

As regards utterance environment, the word-final or coda environment for /t/ presents itself as a site of notable variation, particularly favourable to phenomena such as glottalization. Future work could consider the interaction between voice quality and the coda /t/ environment with this sample [27]. Our preliminary observations show that the children exhibit creaky voice, breathy voice and whispery voice. We also plan to examine VOT and voice termination time (VTT) [see 28] in the future, and to investigate the effects of word frequency and individual differences. Previous work has found adult early simultaneous Spanish-English bilinguals to have mastered VOT in English (as compared to late bilinguals with “compromise” VOT values [17]; see also [16]). We would like to provide a first indication that simultaneous bilingual children acquire the ability to inhibit the non-target language from an early age and more effectively than sequential bilinguals, as proposed by the L2LP model [29, 30].

5. Acknowledgements

Supported by Australian Research Council grants to Author 3 (CE140100041, FT160100514) and Author 1 and Author 3 (LP210300631). With thanks to participating children and parents, Deeahn Sako for data collection, and Minh Anh Tran and Linh Tran for coding assistance.

6. References

[1] D. Loakes and K. McDougall, "Individual Variation in the Frication of Voiceless Plosives in Australian English: A Study of Twins' Speech," *Australian Journal of Linguistics,* vol. 30, no. 2, pp. 155-181, 2010.

[2] D. Loakes, K. McDougall, and A. Gregory, "Variation in /t/ in Aboriginal and Mainstream Australian Englishes," in *Australasian Speech Science and Technology Conference*, Canberra, R. Billington, Ed., 2022: ASSTA, pp. 61-65.

[3] L. Tollfree, "Variation and change in Australian consonants: reduction of /t/," in *Varieties of English Around the World: English in Australia*, D. Blair and P. Collins Eds. Amsterdam: John Benjamins, 2001, pp. 45-67.

[4] B. Horvath, *Variation in Australian English: The Sociolects of Sydney*. Cambridge; New York: Cambridge University Press, 1985.

[5] P. Foulkes, G. J. Docherty, and D. J. L. Watt, "The emergence of structured variation," *University of Pennsylvania Working Papers in Linguistics,* vol. 7, no. 3, pp. 67-84, 2001.

[6] G. Docherty, P. Foulkes, B. Dodd, and L. Milroy, "The emergence of structured variation in the speech of Tyneside infants," Final report to the United Kingdom Economic and Social Research Council, grant R000 237417, 2002.

[7] J. Roberts, "As old becomes new: Glottalization in Vermont," *American Speech,* vol. 81, no. 3, pp. 227-249, 2006.

[8] J. Milroy, L. Milroy, S. Hartley, and D. Walshaw, "Glottal stops and Tyneside glottalization: Competing patterns of variation and change in British English," *Language Variation and Change,* vol. 6, no. 3, pp. 327-357, 1994.

[9] C. Tait and M. Tabain, "Patterns of gender variation in the speech of primary school-aged children in Australian English: the case of /p t k/," in *Australasian Speech Science and Technology Conference*, Parramatta, 2016: ASSTA, pp. 65-68.

[10] J. C. L. Ingram, "Connected speech processes in Australian English," *Australian Journal of Linguistics,* vol. 9, no. 1, pp. 21-49, 1989.

[11] D. Loakes, K. McDougall, J. Clothier, J. Hajek, and J. Fletcher, "Sociophonetic variability of post-vocalic /t/ in Aboriginal and mainstream Australian English," in *Australasian Speech Science and Technology Conference*, Sydney, 2018: ASSTA, pp. 5-8.

[12] J. Penney, F. Cox, and A. Gibson, "Hiatus resolution and linguistic diversity in Australian English," *Phonetica,* vol. 81, no. 2, pp. 119-152, 2024.

[13] A. Gibson, F. Cox, and J. Penney, "Acquiring allophony: GOOSE and SCHOOL vowels in the speech of Australian children," in *Proceedings of the 20th International Congress of Phonetic Sciences*, Prague, Czechia, R. Skarnitzl and J. Volín, Eds., 2023: Guarant International, pp. 3750-3754.

[14] Australian Bureau of Statistics (ABS), "Census of Population and Housing," 2021. Accessed: 18 June 2024. [Online]. Available: https://profile.id.com.au/australia/language

[15] C. E. Gildersleeve-Neumann, E. S. Kester, B. L. Davis, and E. D. Peña, "English speech sound development in preschool-aged children from bilingual English-Spanish environments," *Language, Speech and Hearing Services in Schools,* vol. 39, no. 3, pp. 314-328, 2008.

[16] D. F. Thornburgh and J. H. Ryalls, "Voice onset time in Spanish-English bilinguals: Early versus late learners of English," *Journal of Communication Disorders,* vol. 31, no. 3, pp. 215-229, 1998.

[17] J. E. Flege, "Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language," *Journal of the Acoustical Society of America,* vol. 89, no. 1, pp. 395-411, 1991.

[18] P. Escudero, G. Pino Escobar, M. Hernandez Gallego, and C. Diskin-Holdaway, "The Little Multilingual Minds corpus: Educators' and children's speech in heritage languages," presented at the Workshop on community language corpora in Australia, Canberra, 2023.

[19] 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 2024, doi: doi.org/10.31219/osf.io/rvjcg.

[20] 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: doi.org/10.31219/osf.io/8e4nf.

[21] P. Escudero, G. Pino Escobar, C. G. Casey, and K. Sommer, "Four-Year-Old's Online Versus Face-to-Face Word Learning via eBooks," *Frontiers in Psychology,* vol. 12, 2021, Art no. 610975.

[22] 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*, 2023: PMLR, pp. 28492-28518.

[23] T. Kisler, U. D. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language,* vol. 45, pp. 326–347, 2017.

[24] *R: A Language and Environment for Statistical Computing*. (2024). R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: https://www.R-project.org

[25] B. McMurray, K. A. Kovack-Lesh, D. Goodwin, and W. McEchron, "Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence?," *Cognition,* vol. 129, no. 2, pp. 362-378, 2013.

[26] M. J. Jones and K. McDougall, "A comparative acoustic study of Australian English fricated /t/: assessing the Irish (English) link," in *Australasian Speech Science and Technology Conference*, Auckland, P. Warren and C. I. Watson, Eds., 2006: ASSTA.

[27] J. Penney, F. Cox, and A. Szakay, "Glottalisation, coda voicing, and phrase position in Australian English," *Journal of the Acoustical Society of America,* vol. 148, 2020, Art no. 3232.

[28] R. Mailhammer, S. Sherwood, and H. Stoakes, "The inconspicuous substratum: Indigenous Australian languages and the phonetics of stop contrasts in English on Croker Island," *English World-Wide,* vol. 41, no. 2, pp. 162-192, 2020.

[29] P. Escudero, *Linguistic Perception and Second Language Acquisition: Explaining the attainment of optimal phonological categorization.* Utrecht University: LOT Dissertation Series 113, 2005.

[30] P. Escudero and K. Yazawa, "The Second Language Linguistic Perception Model (L2LP)," in *The Cambridge Handbook of Bilingual Phonetics and Phonology*, M. Amengual Ed. Cambridge: CUP, in press.