BUCLD 46 Proceedings
To be published in 2022 by Cascadilla Press
Rights forms signed by all authors

Sensitivity to Fine-Grained Phonetic Details in Childhood: Differences between First- and Second-Language Learners

Félix Desmeules-Trudel, Elizabeth K. Johnson, and Craig G. Chambers*

1. Introduction

Speech perception has traditionally been described as a "categorical" phenomenon whereby, even early in development, sounds that vary along an acoustic continuum are perceptually classified as members of discrete phonological categories according to the structure of the listeners' phonological systems (Harnad, 2003; Kuhl, 1992; 1994; Kuhl et al., 2008; Liberman et al., 1957). However, recent work in psycholinguistics with adult listeners has shown gradient sensitivities, where stepwise variation along an acoustic continuum like voice onset time (VOT) or duration of vowel nasalization has corresponding stepwise effects on the time course of linking unfolding word forms with meanings (e.g., pictures in a Visual World Paradigm; Desmeules-Trudel and Zamuner, 2019; in press; McMurray, Tanenhaus, & Aslin, 2002; McMurray et al., 2008).

The findings from these adult studies suggest that word recognition can be a gradient process in a first language (i.e., in contrast to traditional categorical view of speech perception). However, understanding second language (L2) word recognition is important to model language processing as a whole, given the high prevalence of bilinguals and L2 learners around the world. The contribution of variable phonetic cues to L2 word recognition has not been as thoroughly explored as in the first language (L1), although emerging evidence also seems to point towards sensitivity to variable acoustic cues in an L2 and gradient word recognition abilities. External variables such as age of L2 acquisition also seem to impact L2 listeners' ability (Desmeules-Trudel and Zamuner, in press). For

* Félix Desmeules-Trudel, University of Toronto Mississauga, felix.desmeules.trudel@utoronto.ca; Elizabeth K. Johnson, University of Toronto Mississauga; Craig G. Chambers, University of Toronto Mississauga. This research was funded by a University of Toronto Mississauga Postdoctoral Fellowship award to the first author. We are grateful to members of the CLASS and PAL labs, Dr. Jessamyn Schertz and members of the PCL cluster at UTM for their feedback on study design, and to Kaitlyn Harris for her contribution to recruitment and data collection.

instance, patterns of recognition of early L2 (English-L1) learners (before 5 years old) of Canadian French were impacted by within-category variability in a way resembling that of native French listeners, using the nasal vowel contrast which exists in their L2 (e.g., *pain* [pɛ̃] 'bread' and *paix* [pɛ] 'peace'). However, late learners who had learned French in a mostly-classroom setting displayed more-categorical patterns of word recognition, suggesting that they were not as sensitive to perceiving variability in vowel nasalization duration (Desmeules-Trudel and Zamuner, in press).

The question of sensitivity to phonetic details and word recognition patterns has not yet been explored in the context of L2 learning across development. This is important for theories of speech processing because of the ever-present phonetic variability in the environment as children learn and process language information. Furthermore, given the sheer number of children who learn an L2 on the way to adolescence and adulthood, advancing our understanding of these phenomena helps promote inclusivity in psycholinguistic inquiry.

In the language development literature, it has been shown that English-speaking monolingual toddlers are sensitive to mispronunciations that vary in the degree of mismatch with the target (White and Morgan, 2008) and to degrees of coarticulatory shading (Desmeules-Trudel, Moore, & Zamuner, 2020; Mahr et al., 2015; Paquette-Smith, Fecher, & Johnson, 2016; Zamuner, Moore, & Desmeules-Trudel, 2016) in a way that is simlar to native-English adults (Beddor et al., 2013; Dahan et al., 2001). This suggests that toddlers and older children up to 6 years old are sensitive to fine-grained phonetic details, and, like adults, are likely to be able to use acoustic variations when recognizing words (White and Morgan, 2008). Because L2 learning is well-known to have impacts on schooling outcomes (Bialystok, 2018) and information processing across the life span (Vinerte and Sabourin, 2019), we explored children's sensitivity to variations in vowel nasalization, directly comparing L2 learners and L1 listeners in the same task. We investigated whether English-speaking five-year-olds, who have relatively robust L1 categories, are sensitive to within-category variability in newly-learned French-like words. We also tested native French-speaking age-matched listeners, in order to determine their word recognition patterns in their L1. This design allows us to compare whether children display gradient or categorical patterns of recognition when presented with variable word forms in their L1 and an L2.

Here we focus on nasalization. In French, nasal vowels (e.g., *bain* [bɛ̃] 'bathtub') are extensively nasalized and contrast with oral vowels (which can nonetheless be phonetically nasalized when preceding nasal consonants, e.g., *beigne* [bɛɲ] 'doughnut'; Delvaux et al., 2008; Desmeules-Trudel and Brunelle, 2018). In English, nasalized vowels do not contrast with oral vowels (Beddor, 2009), but vowels can be phonetically nasalized when they are followed by a nasal consonant: e.g., [æ] in the word *can* is phonetically nasalized [kæ̃n]. The adult work on word recognition using vowel nasalization has shown that English-speaking adults and children can use non-contrastive nasalization in their L1 during word recognition. In contrast, in French, although nasalization can be used in a gradient manner during word recognition by L1 listeners (Desmeules-Trudel and Zamuner, 2019), L2 French learners show a different pattern depending on age of acquisition. Whereas patterns are gradient-like for early learners, they are more categorical for late learners (Desmeules-Trudel and Zamuner, in press).

If children are like adults, French-speaking children should display gradient patterns of word recognition and English-speaking children (with no knowledge of French) should display more-categorical patterns of word recognition. The latter patterns could reflect young learners' inability to discriminate fine-grained acoustic cues associated with the target vowels (Best and Tyler, 2007). This would mean that even at an early age, children are highly sensitive to acoustic variability in their L1, but that the structure of the phonological system, even if not fully mature, would impact word recognition in an L2. Alternatively, all children may show categorical patterns of word recognition whereby they would not yet display sensitivity to fine-grained acoustic (i.e., non-contrastive) variations. This would mean that the use of fine-grained phonetic details in word recognition emerges comparatively late in development.

We explore these questions using a methodology where we used a word learning task followed by a word recognition task, with stimuli that contained the French oral-nasal contrast for vowels, which is absent from English.

2. Methods

2.1 Participants

We recruited two groups of children in the current study: a group of English-speaking children and a group of French-speaking children. The English-speaking group included 24 (Canadian) English-speaking children. Their age ranged from 4;5 to 5;11 years old ($\bar{x}$ = 5;2) and they had not been exposed for French or any other language containing nasal vowels (under 5% of the time, according to parental report on a background language questionnaire). We also recruited 21 French-speaking children (a mix of European and Canadian French parents, all children lived in Canada, mostly the province of Québec, the Greater Toronto Area, and Southern Ontario), ranging from 4;6 to 6;3 years old ($\bar{x}$ = 5;5). According to parental report on a background language questionnaire, all children in the latter group were French-dominant native speakers of French, but it is important to note that some children were bilingual in English to various degrees since they lived in an English-majority setting.

2.2 Stimuli

Stimuli were composed of two minimal pairs of bisyllabic, French-like nonwords containing a final nasal vowel, i.e., [dɔpɛ̃], [ɡubã], or a final complex nucleus with an oral vowel and nasal consonant, i.e., [dɔpɛn], [ɡuban]. Each nonword was associated with a picture of a friendly alien-like creature (Figure 1).

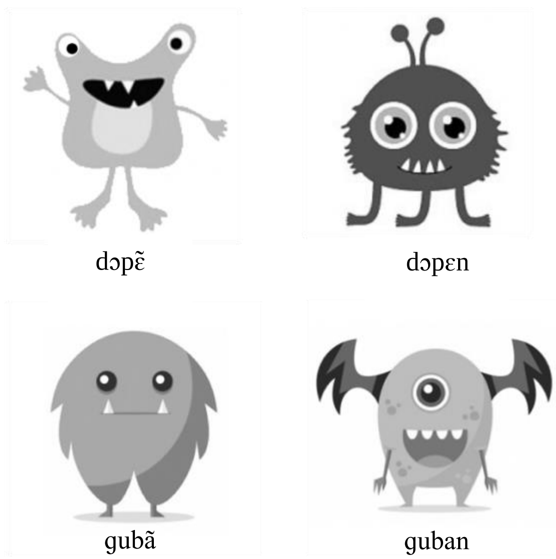

Figure 1. International phonetic alphabet transcriptions of the auditory stimuli and associated pictures (coloured in the actual experiment) for each nonword.

Auditory stimuli were recorded by a male native speaker of Canadian French. Individual tokens' spectrograms of each word were inspected, and tokens were selected based on (visual) similarity in formant values (F1 and F2) between the nasal and oral (final) vowel, as observed by a trained phonetician (first author). This operation was necessary in order to proceed with the stimulus cross-splicing, aiming to ensure smooth transitions at the cross-splicing point, free of clicks or undesired noise in the recordings (for similar cross-splicing procedures, see Desmeules-Trudel et al., 2020, and Desmeules-Trudel and Zamuner, in press).

The vowels of the selected tokens were cross-spliced into another token of the target stimulus that included the initial syllable and onset consonant of the second syllable (as well as coda consonants in the case of oral vowels). In other words, a vowel from a selected token replaced the vowel from another token that was similar in formant values to control for the (potential) effect of cross-splicing. That process yielded four cross-spliced stimuli with vowels in a 0N condition that were not nasalized (i.e., they were oral for their entire duration; [dɔpɛ̱n] and [gub<u>a</u>n]), and vowels in a 100N condition that were entirely nasalized (i.e., they were nasal for their entire duration; [dɔp<u>ɛ̃</u>] and [gub<u>ã</u>]). A third set of cross-spliced stimuli was engineered to provide an intermediate nasalization duration condition: 50N. In this condition, 50% duration of an (selected) oral vowel was followed by 50% duration of a (selected) nasal vowel. The final nasal consonant was shortened to 50 ms. In other words, 50N stimuli contained a final vowel that was nasalized for 50% of its duration towards offset, followed by a short nasal consonant. The 0N and 100N stimuli were presented to the listeners in the learning task, and stimuli in all three conditions (0N, 50N and 100N) were presented in the word recognition task. Sample spectrograms for each condition are displayed in Figure 2.

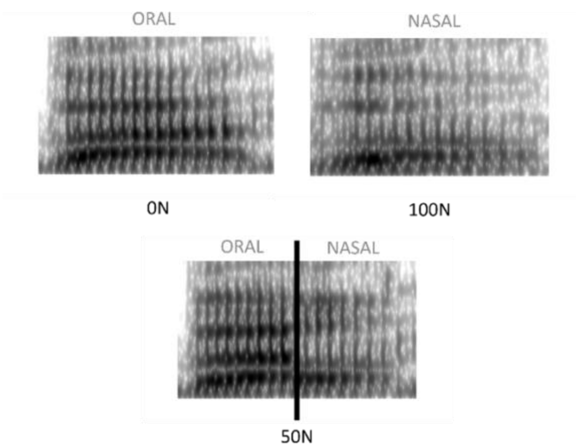

Figure 2. Sample spectrograms of the [a]-[ã] vowels in the 0N, 100N and 50N conditions. The dark black line on the 50N spectrogram represents the point of transition between the oral (left) and nasal (right) portions of the vowel.

3. Learning task

As mentioned above, all children first completed a word learning task. During the procedure, which was implemented online using Gorilla (Anwyl-Irvine et al., 2020), children were first presented with a single target picture in the middle of a display (either a laptop, home computer or tablet, in their home). After 2 seconds of silence, the corresponding auditory stimulus played (see Figure 1 above for the image-word correspondences). Children completed four trials of this "passive learning" phase, i.e., one for each word-image association.

After the passive learning phase, all children completed an "active learning" phase, in which they saw two images corresponding to a minimal pair appearing on their display for 2 seconds, followed by an auditory stimulus corresponding to one of the two pictures (0N or 100N stimuli). They were then asked to guess which of the two pictures corresponded to the auditory stimulus and received correct or incorrect feedback (“Good job!” accompanied by a green checkmark on the display, or “Oops, that was the other one!” with a red X on the display). Children heard each auditory stimulus 8 times, for a total of 32 learning trials overall, divided in 4 blocks of eight trials. Blocks were separated by cartoon animations of a spaceship travelling through space, fueled by the child’s correct answers, to maintain motivation to participate and complete the task. The task lasted between 5 and 7 minutes, depending on the children’s response speed.

To assess children’s learning abilities throughout the task, we compiled the number of correct responses per block, converted into rates of correct responses. Results are presented in Figure 3, and show that, by the third block (English-speaking) and fourth block (French-speaking) children had rates of correct responses that were significantly above chance (more than 4/8 correct responses), as determined by one-tailed *t*-tests (English-speaking children had rates of correct responses that were not different from chance in the 4th block, which might be attributable to noise in the data). This suggests that both groups were able to learn the target words, regardless of the language background. Interestingly, French-speaking children had overall higher rates of correct responses across all blocks, as observed on Figure 3. This trend can be interpreted as a manifestation of their phonological system facilitating learning of words that contain a nasal contrast, (i.e., they have already developed categories for these speech sounds), as opposed to English-speaking children whose phonological system does not require nasal and non-nasal vowels to be differentiated.

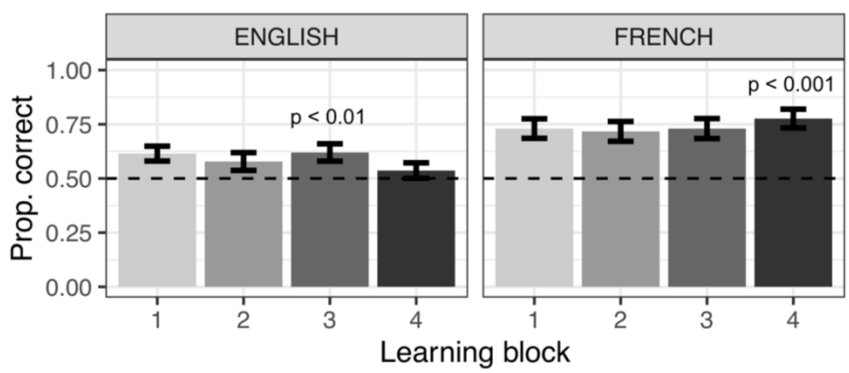

Figure 3. Rates of correct responses in the learning task per block for English-speaking and French-speaking children.

4. Word recognition task

Following learning, all children completed a word recognition task to assess their sensitivity to variations in duration of nasalization. The recognition task was also implemented online using Gorilla (Anwyl-Irvine et al., 2020). For this task, which resembled the active learning phase of the preceding task, children were presented with three pictures of aliens– two members of a minimal pair, and a never-seen-before picture of another creature. We included a third picture on the display because, in addition to hearing 0N and 100N stimuli, children also heard 50N auditory stimuli that they had not encountered in the learning phase. Thus, the third creature provided a meaningful choice in the case children interpreted the 50N stimuli as a different object label and not merely a poor token of a pre-learned label. After 2 seconds of silence, the corresponding auditory stimulus was played (see Figure 1 above for the image-word correspondences), after which children were asked to choose which picture they thought corresponded to the label that was heard. Children were not provided with feedback during the recognition phase in order to minimize additional learning.

Listeners completed 24 recognition trials, divided into four blocks of six trials. Within one block, each auditory stimulus was presented once ([dɔpɛ̃] - [dɔpɛn] in the 0N, 50N and 100N conditions, and [gubã]-[guban] in the 0N, 50N and 100N conditions). Block were separated by cartoon animations of a friendly alien creature distributing stars to children who had to complete a board of four stars to succeed through the task. These were included to maintain interest and motivation. The word recognition task took approximately 5 minutes to complete.

We compiled the number of times that children selected the picture associated with the 100N (fully nasal) stimuli, for each nasalization duration condition, in English-speaking and French-speaking listeners (Figure 4), which were further converted in rates of 100N responses. A binomial linear mixed-effects model (Table 1; individual participants as random effect), implemented in R version 3.6.2 (R Core Team, 2019) with lme4 version 1.1-21 (Bates et al., 2015), revealed that for the English-speaking group, rates of 100N responses did not significantly differ between 50N and 0N stimuli. However, the difference between the 100N response rates in the 50N and 100N was significant, suggesting a categorical pattern of newly-learned word recognition in this group, similar to L2 adults (Desmeules-Trudel and Zamuner, in press).

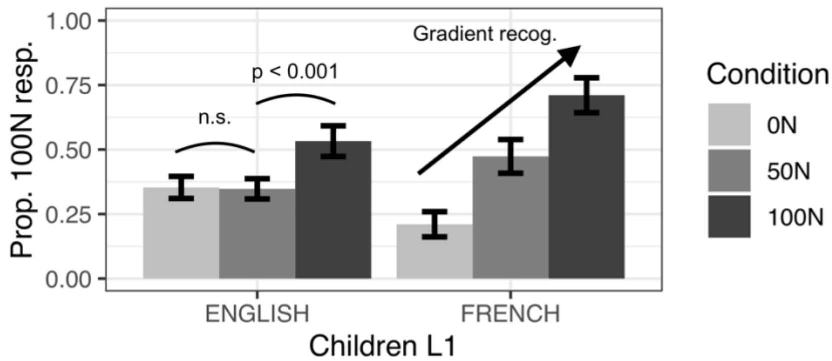

Figure 4. Rates of 100N responses in the word recognition task for each nasalization duration condition, for English-speaking and French-speaking children.

In contrast, French-speaking children showed a distinct pattern of results. Another binomial mixed-effects model (Table 2; individual participants as random effect) revealed that the rates of 100N responses were significantly lower for the 0N than for the 50N conditions, and significantly higher in the 100N condition than in the 50N condition for the French-speaking children. This suggests a gradient pattern of word recognition, similar to L1 adults (Desmeules-Trudel and Zamuner, 2019).

Table 1. Results of the linear mixed-effects model on rate of 100N responses in the recognition task for the English-speaking group.

Formula: resp100N ~ condition + (1|participant), data=recog_eng, family=binomial				
	Estimate	Std. Error	*Z* value	*p* value
Intercept (50N)	-0.76	0.21	-3.68	<0.001***
50N vs. 0N	0.03	0.22	0.11	0.91
50N vs. 100N	0.8	0.22	3.63	<0.001***

Table 2. Results of the linear mixed-effects model on rate of 100N responses in the recognition task for the French-speaking group.

Formula: resp100N ~ condition + (1|participant), data=recog_fr, family=binomial				
	Estimate	Std. Error	*Z* value	*p* value
Intercept (50N)	-0.08	0.19	-0.42	0.68
50N vs. 0N	-1.36	-.26	-5.25	<0.001***
50N vs. 100N	1.09	0.24	4.47	<0.001***

5. Discussion

In this study, we investigated whether 4- to 6-year-olds were sensitive to fine-grained phonetic details in the first (L1) and second language (L2) during word recognition. Indeed, the variability children are presented with in the input when acquiring one or more languages can pose difficulties for learning and processing words. In order to better understand how this kind of variability impacts L1 and L2 learning and processing, participants were asked to learn new words in their L1 and an L2, and their learning was then tested in a subsequent recognition task. Words contained either an oral or a nasal vowel, which are phonologically contrastive in French but not in English. Patterns found in the adult literature led to the prediction that English-speaking children should display categorical patterns of recognition, whereas French-speaking children should display gradient patterns of recognition (Desmeules-Trudel and Zamuner, 2019; in press).

Both groups of children successfully completed the (non)word learning task, in which they had to link images of novel creatures to new auditory words. Through repetitions and feedback on the accuracy of their responses, both groups were able to learn the four target words at levels significantly above chance, even in an L2 that they were not familiar with prior to the study. Interestingly, L1 children had higher overall rates of correct responses throughout the task, suggesting that the pre-established category in their phonological system facilitated word learning. These results underscore the capacity for children to quickly grasp new linguistic information, as well as the point that experience with phonological contrasts facilitates word learning in the L1.

Following learning, all children completed a word recognition task. Auditory stimuli varied in duration of nasalization along a continuum: vowels were nasalized for 0% of their duration, 50% of their duration, or 100% of their duration. Results showed that English-speaking children displayed a categorical pattern of word recognition for the non-native sound contrasts, suggesting that they were sensitive to vowel nasalization to some extent, but that this sensitivity was limited to the "extremes" of the continuum. In other words, the structure of their native phonological system, developed on the basis of input that did not require them to significantly contrast oral and nasal vowels, impacted the way they used nasalized vowels for word recognition.

Inversely, French-speaking children displayed a gradient pattern of word recognition for the same stimuli (i.e., the proportion of word choice that corresponded to fully nasalized stimuli gradually increased as the physical duration of nasalization increased on the stimuli). This shows that they were greatly sensitive to acoustic variability in the L1 and that they were able to use this variability in a gradient fashion, despite the categorical distinction between oral and nasal vowels in their L1. This is similar to what has been documented in adults using vowel nasalization (Desmeules-Trudel & Zamuner, 2019; in press) as well as children in their L1 (White and Morgan, 2008), and might be due to greater sensitivity to the difference in nasalization duration between oral and nasal vowels in the French native input they received during acquisition.

The current study constitutes an important step towards a better understanding of word processing abilities for non-monolingual listeners. We showed that children's learning abilities can rapidly translate to performance on a word recognition task, and that children were sensitive to fine-grained phonetic details in their L2, to some extent. Indeed, although English-speaking children displayed a categorical pattern of word recognition in their L2, they were nonetheless able to differentiate stimuli that were nasalized for 50% of their duration from fully nasalized stimuli. This again underscores children's abilities to quickly learn new information and brings further insights into the way the L1 phonological system impacts sound categorization in both the L1 and the L2.

It is noteworthy that the results we report here were obtained using an all-online procedure, with children completing all steps of the study in their home. Not only were we able to conduct a relatively complex task remotely, but the results that we obtained are meaningful and fit well within the literature on child and L2 speech perception/word recognition. Thus, the current study bodes well for future online studies on fine-grained phonetic processing, and adds to a growing body of literature advocating the inclusion of more diverse populations in psycholinguistic studies (Bacon, Weaser, & Saffran, 2021), something that is often difficult to do in a lab setting. Future directions could include in-person studies using eye tracking in the Visual World Paradigm, which would facilitate comparisons with other studies and enable us to obtain a finer picture of the mechanisms behind L2 word processing in children by considering the dynamics of unfolding speech. Such experiments would also enable us to further our understanding of speech processing and learning, in turn extending models to bilingual and multilingual learners and gaining a more precise understanding of the developing mind.

References

Anwyl-Irvine, Alexander L., Massonnié, Jessica, Flitton, Adam, Kirkham, Natasha, & Evershed, Jo K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388-407.

Bacon, Desia, Weaser, Haley, & Saffran, Jenny. (2021). A framework for online experimenter-moderated looking-time studies assessing infants' linguistic knowledge. *Frontiers in Psychology*, *12*, Article 703839.

Bates, Douglas, Maechler, Martin, Bolker, Ben, & Walker, Steve. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Beddor, Patrice S. (2009). A coarticulatory path to sound change. *Language*, *85*(4), 785-821.

Beddor, Patrice S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, *133*(4), 2350–2366.

Best, Catherine T., & Tyler, Michael D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp. 13-34). Amsterdam, NL: John Benjamins.

Bialystok, Ellen. (2018). Bilingual education for young children: Review of the effects and consequences. *International Journal of Bilingual Education and Bilingualism*, *21*(6), 666-679.

Dahan, Delphine, Magnuson, James S., Tanenhaus, Michael K., & Hogan, Ellen M. (2001) Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507-534.

Delvaux, Véronique, Demolin, Didier, Harmegniew, Bernard, & Soquet, Alain. (2008). The aerodynamics of nasalization in French. *Journal of Phonetics*, *36*, 578–606.

Desmeules-Trudel, Félix, & Brunelle, Marc. (2018). Phonotactic restrictions condition the realization of vowel nasality and nasal coarticulation: Duration and airflow measurements in Québécois French and Brazilian Portuguese. *Journal of Phonetics*, *69*, 43-61.

Desmeules-Trudel, Félix, Moore, Charlotte E., & Zamuner, Tania S. (2020). Monolingual and bilingual children's processing of coarticulation cues during spoken word recognition. *Journal of Child Language*, *47*, 1189-1206.

Desmeules-Trudel, Félix, & Zamuner, Tania S. (2019). Gradient and categorical patterns of spoken word recognition and processing of phonetic details. *Attention, Perception, & Psychophysics*, *81*(5), 1654-1672.

Desmeules-Trudel, Félix, & Zamuner, Tania S. (in press). Spoken word recognition in a second language: The importance of phonetic details. *Second Language Research*, Online First, 1-30.

Harnad, Stevan. (2003). Categorical perception. In *Encyclopedia of Cognitive Science* (pp. 448–452). Malden, MA: Wiley-Blackwell.

Kuhl, Patricia K. (1992). Psychoacoustics and speech perception: Internal standards, perceptual anchors, and prototypes. In L. A. Werner & E. W. Rubel (Eds.), *Developmental Psychoacoustics* (pp. 293–332). Washington, DC: American Psychological Association.

Kuhl, Patricia K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, *4*(6), 812–822.

Kuhl, Patricia K., Conboy, Barbara T., Coffey-Corina, Sharon, Padden, Denise, Rivera-Gaxiola, Maritza, & Nelson, Tobey. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979–1000.

Liberman, Aalvin M., Harris, Katherine S., Hoffman, Howard S., & Griffith, Belver C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368.

Mahr, Tristan, McMillan, Brianna T. M., Saffran, Jenny R., Ellis Weismer, Susan, & Edwards, Jan. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, *142*, 345–350.

McMurray, Bob, Clayards, Meghan A., Tanenhaus, Michael K., & Aslin, Richard N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064–1071.

McMurray, Bob, Tanenhaus, Michael K., & Aslin, Richard N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42.

Paquette-Smith, Melissa, Fecher, Natalie, & Johnson, Elizabeth K. (2016). Two-year-olds' sensitivity to subphonemic mismatch during online spoken word recognition. *Attention, Perception, & Psychophysics*, *78*(8), 2329–2340.

R Core Team (2019) *R: A language and environment for statistical computing* [software]. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org (accessed June 2021).

Vīnerte, Santa, & Sabourin, Laura. (2019). Reviewing the bilingual cognitive control literature: Can a brain-based approach resolve the debate? *Canadian Journal of Experimental Psychology/Revue Canadienne de psychologie expérimentale*, *73*(2), 118-134.

White, Katherine S., & Morgan, James L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*, 114-132.

Zamuner, Tania S., Moore, Charlotte E., & Desmeules-Trudel, Félix. (2016). Toddlers' sensitivity to within-word coarticulation during spoken word recognition: Developmental differences in lexical competition. *Journal of Experimental Child Psychology*, *152*, 136–148.