



**HAL**  
open science

## Variation in phonological bias: Bias for vowels, rather than consonants or tones in lexical processing by Cantonese-learning toddlers

Hui Chen, Daniel T Lee, Zili Luo, Regine y Lai, Hintat Cheung, Thierry Nazzi

### ► To cite this version:

Hui Chen, Daniel T Lee, Zili Luo, Regine y Lai, Hintat Cheung, et al.. Variation in phonological bias: Bias for vowels, rather than consonants or tones in lexical processing by Cantonese-learning toddlers. *Cognition*, 2021, 213, pp.104486. 10.1016/j.cognition.2020.104486 . hal-03391035

**HAL Id: hal-03391035**

**<https://hal.science/hal-03391035v1>**

Submitted on 21 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Variation in phonological bias: Bias for vowels, rather than consonants or tones in lexical processing by Cantonese-learning toddlers

Hui Chen<sup>1</sup>, Daniel T. Lee<sup>2</sup>, Zili Luo<sup>2</sup>, Regine Y. Lai<sup>3</sup>, Hintat Cheung<sup>2</sup>, & Thierry Nazzi<sup>1</sup>

<sup>1</sup> Integrative Neuroscience and Cognition Center, CNRS & Université Paris Descartes

<sup>2</sup> The Education University of Hong Kong

<sup>3</sup> The Chinese University of Hong Kong

Declarations of interest: none

Acknowledgments

This work was funded by a joint ANR-RGC grant (A-HKIEd801/15 to H Cheung and ANR-15-CE28-0011 to T Nazzi). We thank all the parents and infants who participated in the study. In memoriam Will Bishop 06.10.1971-05.12.2019.

Address for Correspondence

Thierry Nazzi or Hui Chen. Integrative Neuroscience and Cognition Center (INCC), Language and Cognition Team. CNRS - Université de Paris. Institut Pluridisciplinaire des Saints Pères. 45 rue des Sts Pères. 75270 Paris cedex 06. e-mail : thierry.nazzi@parisdescartes.fr; joan\_chen19@hotmail.com

## Abstract

Consonants and vowels have been considered to fulfill different functions in language processing, vowels being more important for prosodic and syntactic processes and consonants for lexically related processes (Nespor et al., 2003). This C-bias hypothesis in lexical processing is supported by studies with adults and infants in many languages such as English, French, Spanish, although a few studies, on Danish and Mandarin, suggest the existence of cross-linguistic variation. The present study explores whether a C-bias exists in a tone language with a complex tone system, Cantonese, by comparing the relative weight given to consonants, vowels, and also tones during word learning. To do so, looking behaviors of Cantonese-learning 20- and 30-month-olds (24 children per age/condition, 6 groups) were recorded by an eyetracker while they watched animated cartoons in Cantonese to learn pairs of novel words. The words differed minimally by either a consonant (e.g., /tɕɛ6/ vs. /kɕɛ6/), a vowel (e.g., /k<sup>h</sup>im3/ vs. /k<sup>h</sup>ɛm3/), or a tone (e.g., T2 vs. T5). Analyses on proportional looking times revealed significant learning in 30-month-olds only, and at that age, only for the vowel contrasts. Growth curve analyses revealed better performance for the vowel condition compared to the other two conditions. The present findings establish a V-bias in Cantonese-learning 30-month-olds, adding new evidence from that tone language that the C-bias in lexical processing is not language-general. Implications for theoretical discussions on the origins of this phonological bias, and the impact of tones in early language acquisition, are discussed.

Keywords: phonological bias, word learning, minimal pairs, Cantonese

## 1. Introduction

Every human language is a remarkably sophisticated and highly complex system, which is yet acquired quickly by most human children. This rapid language acquisition has been shown to be based on powerful processing abilities that are already present at birth, allowing perceptual discrimination of linguistic sounds at the segmental (e.g., Eimas, Siqueland, Jusczyk & Vigorito, 1971; Mahmoudzadeh, Dehaene-Lambertz, Fournier, Kongolo, Goudjil, Dubois, Grebe & Wallois, 2013) and prosodic (lexical stress: Sansavini, Bertoncini & Giovanelli, 1997; pitch contour: Nazzi, Floccia & Bertoncini, 1998; language rhythm: Nazzi, Bertoncini & Mehler, 1998) levels by newborn infants. Some of those abilities are probably set into place in the last semester of gestation, allowing for some influences of the environmental language already at birth, at least for abilities relying on prosodic information (preference for mother's voice, Mehler, Bertoncini, Barriere & Jassik-Gerschenfeld, 1978; DeCasper & Fifer 1980; recognition of maternal language rhythm: Mehler et al., 1988, Moon, Panneton-Cooper & Fifer, 1993; memorization of short stories: DeCasper & Spence, 1986; language-specific modulation of rhythmic grouping: Abboub, Nazzi & Gervain, 2016). Further acquisitions happen in the first year of life, and include the repertoire of native phonemes (Werker & Tees, 1984; Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992), native prosodic properties (word level: Jusczyk, Cutler & Redanz, 1993; Hoehle, Bijeljac-Babic, Herold, Weissenborn & Nazzi, 2009; higher levels: van Ommen, Boll-Avetisyan, Larraza, Wellmann, Bijeljac-Babic, Höhle & Nazzi, 2020), and the acquisition of lexical tones in tone languages (Mattock & Burnham, 2006; Yeung, Chen & Werker, 2013). Comprehension of some words frequently used in the infants' environment starts by 6 months, months before they start being produced (Tincoff & Jusczyk, 1999, 2012; Bergelson & Swingley, 2012). Moreover, very early on during language acquisition, links between the acquisition of the different linguistic levels have been found (Stager & Werker, 1997; Newman, Ratner, Jusczyk, Jusczyk & Dow, 2006; DePaolis, Vihman & Keren-Portnoy, 2011). Here, we focus on a proposed link between phonological and lexical acquisition, namely the fact that some phonological elements have been found to be more important at the lexical level than others (Nespor, Peña & Mehler, 2003). However, this link appears to be linguistically variable, and to be acquired during development. The present study explores this phenomenon in a language with a complex tone system, Cantonese. It explores the interplay between word

learning and phonological processing (of consonants, vowels and tones) by investigating how Cantonese-learning toddlers learn pairs of Cantonese pseudowords minimally differing by a consonant, a vowel or a tone.

The proposal evaluated in the present study is the “division of labor” hypothesis according to which consonants carry more information about the lexicon, whereas vowels play a more important role in syntactic and prosodic processing (Nespor et al., 2003). More specifically, it explores the bias for consonantal information at the lexical level (henceforward, C-bias). This bias is illustrated by results from word reconstruction studies, that initially found that when asked to transform pseudowords into real words, English-, Dutch-, and Spanish-speaking adults tend to preserve consonantal rather than vocalic information (Cutler, Sebastián-Gallés, Soler- Vilageliu, & van Ooijen, 2000; van Ooijen, 1996). Nazzi and Cutler (2019) reviewed evidence from 38 studies of adult auditory lexical processing, concerning 13 languages from 7 language families and using a variety of lexically-related tasks, which overwhelmingly supports this C-bias in adult lexical processing. However, as they noted, studies on Mandarin and Cantonese, two tone languages, have so far failed to find a C-bias (Wiener & Turnbull, 2016; Wiener, 2020; Gómez, Mok, Ordin, Mehler, & Nespor, 2017; Poltrock, Chen, Kwok, Cheung & Nazzi, 2018), suggesting crosslinguistic variation in the C-bias in adulthood, an issue we return to in more details below.

The review also covered developmental evidence of the C-bias, to which the current study will add. The aim of these studies was to determine the origin of the C-bias found in adulthood, and tested four different proposals. The first one, the innate bias hypothesis, postulates that infants start processing consonants and vowels as distinctive linguistic categories from the onset of language acquisition (Nespor et al., 2003). The other proposals however postulate that the C-bias is learned during acquisition. According to the acoustic/phonetic hypothesis, the C-bias emerges during the first year of life due to consonant/vowel acoustic differences, as consonants are usually shorter, less periodic, and tend to be perceived more categorically than vowels. These acoustic differences would lead to the construction of two phonologically distinct categories in infancy, which would lead to the observation of a phonological consonant bias in toddlers and adults (Floccia, Nazzi, Luche, Poltrock, & Goslin, 2014; Bouchon, Floccia, Fux, Adda- Decker & Nazzi, 2015). In contrast, the lexical hypothesis proposes that the C-bias is a reflection of the linguistic experience of the learners at the lexical

level, from which they will learn that consonants are better, more informative cues to lexical identity than vowels (because more lexical contrasts involve consonants than vowels) once they have acquired a large enough lexicon around 12 months of age (Keidel, Jenison, Kluender & Seidenberg, 2007). Lastly, the phono-lexical hypothesis, acknowledging that phonological and lexical acquisition are intertwined, proposes that the acoustic/phonetic and lexical hypotheses should not be opposed but rather seen as complementary (Nishibayashi & Nazzi, 2016; Poltrock & Nazzi, 2015).

The developmental evidence gathered so far suggests that the C-bias is not innate but acquired, and that its developmental trajectory is language-specific. The first point is supported by studies on French and Italian. In French, there is ample evidence that the C-bias initially found in toddlers (Nazzi, 2005; Havy & Nazzi, 2009) is found as early as 8 to 11 months of age (Nishibayashi & Nazzi, 2016; Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020). However, a reversed V-bias is found in younger French-learning infants, around 5 to 8 months (Bouchon et al., 2015; Nishibayashi & Nazzi, 2016; Von Holzen & Nazzi, 2020). This bias reversal is found even when using the same task at different ages, between 6 and 8 months for word form segmentation (Nishibayashi & Nazzi, 2016), and between 8 and 11 months for own name recognition (Von Holzen & Nazzi, 2020). A similar developmental pattern was found for Italian, in which a V-bias found at birth (Benavides-Varela, Hochmann, Macagno, Nespors, Mehler, 2012) and in 6-month-olds (Hochmann, Benavides-Varela, Nespors, Mehler, & Flo, 2017) gives way to a C-bias by 12 months (Hochmann, Benavides-Varela, Nespors, & Mehler, 2011).

The second point, regarding cross-linguistic variation, is supported by data showing that the early emergence of the C-bias in French and Italian is not found in all languages. Studies on English, exploring lexical recognition or new word learning, have shown the emergence of a C-bias later in development, sometime between 24 and 30 months (Floccia et al., 2014; Nazzi, Floccia, Moquet & Butler, 2009; Mani & Plunkett, 2007, 2010). Moreover, a word learning study with Danish-learning 20-month-olds showed more reliance on vowels than consonants, providing the first evidence of the possibility of a reversed vowel bias (V-bias) in lexical processing in the toddler years (Højen & Nazzi, 2016). It has been suggested that such protracted or reversed biases could stem from acoustic/phonetic properties of these languages, such as the existence of consonant lenition in Danish. Lexical factors might also be implicated, such as a

higher proportion of vowels over consonants in Danish, which might make vocalic skeletons comparatively more informative to identify words.

The present study will contribute new data charting the developmental and crosslinguistic variation in the link between phonological and lexical processing, and finding variation would support the acquired bias hypotheses over the innate bias hypothesis. To do so, it investigates a children population that has not been explored yet, Cantonese-learning children, who were tested at two ages, 20 and 30 months. Tone languages are of interest to further test the C-bias as lexical tones constitute a third type of phonological element on which words can be contrasted, and importantly, they are mostly carried by vowels. This acoustic property might thus affect the relative weight of consonants and vowels, in two opposing ways. One possibility is that tones might make the acoustic realization of vowels more variable in tone than in non-tone languages, making their processing and identification more difficult. The C-bias found in non-tone languages might then be even more pronounced in tone languages. On the contrary, our working hypothesis is that the need for speakers of tone languages to attend to tones to identify words might increase their attention to vowels (which carry them), and thus increase the weight given to vowels compared to consonants in tone languages compared to non-tone languages. This would result in a lack of bias or in a reversed advantage in processing vocalic information. Moreover, syllable structures are very constrained in many tone languages (e.g., no complex onsets and almost no codas), increasing the proportion of vowels within words, which might remove a lexical basis for the emergence of a C-bias.

A few studies have started exploring the issue of whether a C-bias (or a V-bias) is present, both in adulthood and in infancy, in two tone languages, Mandarin and Cantonese. They also compare how tonal information is processed relative to segmental (consonant and vowel) information, the division of labour hypothesis not making assumptions about tonal processing. Two adult studies, using a word reconstruction paradigm based on van Ooijen (1996), compared how Mandarin-speaking adults transform a pseudoword into a real word by changing either a consonant, a vowel (in fact, to conform to Chinese phonology, participants were asked to change the final (rime), corresponding to V, VV, VN or VVN), a tone, or any of the three (Wiener & Turnbull, 2016; Wiener, 2020). Results show effects of condition, corresponding to the fact that these adults preferentially changed tones over both in consonants and vowels/finals (Wiener & Turnbull, 2016), even when removing the stimuli with final nasals and all vowel changes

involving a nasal (Wiener, 2020). Hence, vowels appeared to be the less mutable sound category in Mandarin, contrary to what had been found in non-tone languages, such as Dutch, English and Spanish (van Ooijen, 1996; Cutler et al., 2000), suggesting that more weight is given to vowels in Mandarin. Note however that in these Mandarin experiments (unlike in the experiments on the non-tone languages), the consonant changes always preceded the vowel changes, a positional confound which might have critically contributed to the consonant/vowel asymmetry observed. Yet, the vowel advantage in Mandarin speakers was also found when they processed English stimuli, in which the relative position of consonants and vowels was controlled (Wiener, 2020).

A second study, using an artificial language paradigm, explored whether Cantonese-speaking adults use consonants or vowels (and tones) to segment a fluent speech stream (Gómez et al., 2017). It revealed that they could not use consonantal information alone, but could rely either on vocalic information alone (although the difference between the consonant and vowel conditions was not significant), or more likely on a combination of vocalic and tonal information. These findings support a different balance in the weight given to consonants and vowels by Cantonese-speaking adults as compared to French- or Italian-speaking adults (Bonatti, Peña, Nespor & Mehler, 2005; Toro, Nespor, Mehler & Bonatti, 2008).

A third study used an eye-tracker to evaluate how Cantonese-speaking adults process consonants, vowels and tones when learning new words in their native language (Poltrock et al., 2018). The results failed to find differences in performance between the different conditions, which contrast with the behavior of French-speaking adults, who were found to learn new French words better when the novel words differed in consonants rather than in vowels (Havy, Serres & Nazzi, 2014). These findings might suggest that Cantonese-speaking adults give more weight to vowels than French-speaking adults when learning new words in their native language. However, because Mandarin- and French-speaking adults also failed to show a consonant/vowel asymmetry when learning the Cantonese words in the exact same experimental conditions (Poltrock et al., 2018), much caution is needed in interpreting the Cantonese results. Yet, taken together, these studies on tone languages (Mandarin and Cantonese) suggest that the presence of lexical tones changes the relative weight of consonants and vowels in adult lexical processing, increasing sensitivity to vocalic information. Since the C-bias found in most non-tone languages appears to be set into place early in development, the next section explores what kind of bias is found in children learning a tone language.



In a first study, Singh, Goh & Wewalaarachchi (2015) used a preferential looking paradigm to test bilingual Mandarin-English toddlers (2.5–3.5 years) and preschoolers (4–5 years) on their recognition of familiar Mandarin words, when the words were either correctly pronounced or mispronounced by a consonant, vowel, or tone. Analyses of proportional looking times revealed that the toddlers were more sensitive to tone over both consonant and vowel mispronunciations, while the opposite pattern was found in the preschoolers. Yet, at both ages, no difference was found between sensitivity to consonant versus vowel mispronunciations. This fails to reveal a C- or V-bias in this population, which might result from learning a tone language, although the fact that these children were bilinguals might also have impacted the results.

In a follow-up study, both monolingual Mandarin-learning and bilingual Mandarin-English 24-month-olds were tested on their sensitivity to consonant, vowel and tone mispronunciations of familiar words (Wewalaarachchi, Wong & Singh, 2017). Analyses of proportional looking times revealed that these toddlers were equally sensitive to consonant, vowel and tone mispronunciations, and these effects did not differ across the two linguistic groups, again failing to find a C- or V-bias. But when taking response speed into account, group differences emerged: The monolingual toddlers were most sensitive to tone, then vowel, and then consonant mispronunciations, while the bilingual toddlers were most sensitive to vowel, then consonant, and then tone mispronunciations. While these analyses reveal a V-bias in both groups of 24-month-olds, these effects were only found on the subset of the data in which the toddlers were initially looking at the distractor object when the target was named, which amounted to only 30% of the data, calling for additional, stronger evidence of such a V-bias.

Lastly, monolingual Mandarin-learning and bilingual Mandarin-English 6-year-olds were tested on sensitivity to consonant, vowel and tone mispronunciations of familiar words, using either a preferential looking task or an explicit judgement task (Wewalaarachchi & Singh, 2020). In both experiments, both groups showed least sensitivity to tone mispronunciations, while performance for consonant and vowel mispronunciations did not differ, again failing to provide strong evidence of a phonological bias in Mandarin-learning children.

In summary, the above developmental studies suggest that the C-bias, found in the same age ranges in several non-tone languages, is not present in Mandarin, which supports a different developmental trajectory. This suggests that the presence of lexical tones interferes with the

emergence of a C-bias, rather than reinforcing it. However, the studies failed to provide strong evidence of an opposite V-bias in Mandarin-learning children, while emerging evidence suggests it is present in Mandarin-speaking adults (Wiener & Turnbull, 2016; Wiener, 2019). This calls for further studies to be conducted in Mandarin as well as other tone languages. Here, we explore the presence of a V-bias in Cantonese-learning toddlers. This allows us to test whether the pattern found for Mandarin generalizes to another tone language, thus reinforcing the likelihood that the different trajectory found in Mandarin is a result of the presence of lexical tones. Moreover, we hypothesize that if it is really the presence of tones that induces less reliance on consonants in Mandarin, then stronger evidence of a V-bias might be found in Cantonese compared to Mandarin given its more complex tone system (Cantonese: Bauer & Benedict, 1997; Hashimoto, 1972; Yip, 2002; Mandarin: Cheng, 1966; Duanmu, 2000; Howie, 1976; Wang, 1963). Indeed, there are 6 tones in Cantonese compared to 4 in Mandarin, and identification of tones in Cantonese requires the specification of both their contour (falling, rising...) and their register (pitch height), while only contour is necessary to distinguish Mandarin tones. This difference induces differences in the dimensions that are used by Mandarin and Cantonese-speaking adults to discriminate and identify tones (e.g., Wong, Cheng & Chen, 2018).

Therefore, the present study, which was based on Poltrock et al. (2018), used eyetracking technique to investigate Cantonese-learning 20- and 30-month-olds' ability to learn pairs of Cantonese pseudowords while processing fine phonetic information (consonant versus vowel versus tone information) that is used contrastively in their native language. These ages were chosen because they correspond to either an age, 20 months, at which cross-linguistic differences have been previously found (C-bias in French, no bias in British English but V-bias in Danish), or to the first age, 30 months, at which a C-bias was found in British English (Nazzi, 2005; Nazzi et al., 2009; Floccia et al., 2014; Højen & Nazzi, 2016). Because this was the first study on this population, we opted for a between-participant design, each infant being tested either on consonant, vowel or tone contrasts, in order to obtain more data points (8) per infant in each condition, and thus more reliable data.

On each of the 8 trials, there was a word learning phase, during which, each child saw a pair of cartoons. In each cartoon, an unfamiliar object was presented visually while 6 sentences in Cantonese, each containing a pseudoword labeling that object, were heard. Within each pair, the pseudowords differed by either a consonant, a vowel or a tone, depending on the condition to

which the child was assigned. Following this short learning phase, the children were then tested on whether they had learned the words by presenting them with the two unfamiliar objects side-by-side. To explore the relative strength of the processing of consonant, vowel and tone information, we first analyzed the changes in mean percentage of looking times at the target object before (prenaming phase) and after (postnaming phase) it was named (which also evaluates whether the pseudowords are learned). Second, which motivated our use of an eyetracking procedure rather than a manipulation task as used in many previous studies, we conducted a growth curve analysis on the time course of target looking during the postnaming phase to determine if and when in processing looking times to the target differ between the three conditions. If the presence of lexical tones in a language leads to a V-bias, as suggested by the results on Mandarin, we expect better performance for vowels than for consonants, possibly with faster, sharper responses in the vowel condition. Third, following work by Singh and colleagues on Mandarin-learning children (Singh et al., 2015; Wewalaarachchi et al., 2017; Wewalaarachchi & Singh, 2020), our study will also provide data regarding the processing of tonal information, relative to consonants and vowels, in Cantonese. Fourth, to determine whether word learning performance is linked to vocabulary size, we evaluated the children's performance using a parental report.

## **2. Methods**

### **2.1. Participants**

Seventy-two monolingual Cantonese-learning 20-month-olds (mean age = 20 months 22 days; range: 20 months 1 day – 22 months 0 day; 30 girls, 42 boys) and 72 monolingual Cantonese-learning 30-month-olds were tested and included in the analyses (mean age = 30 months 19 days; range: 30 months 0 day – 31 months 23 days; 34 girls, 38 boys). All toddlers tested were mostly receiving Cantonese input (with some minimal input from another language for some of them, English for the most part), were considered monolinguals by their parents, and both of their parents were native speakers of Cantonese. At each age, 24 toddlers were randomly assigned to one of the 3 groups corresponding to one of the 3 testing conditions: C-contrast condition, V-

contrast condition and T-contrast condition. At each age, across the different conditions, infants were matched for age and vocabulary outcomes/CDI (all  $t_s < 1$ ), and female/male ratio. Thirty additional 20-month-olds were tested and excluded from the analyses due to fussiness/crying ( $N = 16$ ), technical problems ( $N = 1$ ) or because they did not meet the inclusion criteria ( $N = 13$ ; see paragraph on data analysis for details). Twenty-five additional 30-month-olds were tested and excluded from the analyses due to fussiness/crying ( $N = 16$ ) or because they did not meet the inclusion criteria ( $N = 9$ ; see paragraph on data analysis for details).

All toddlers were tested in Hong Kong. Parents of all participants provided written informed consent prior to the experiment. The experiment was approved by the Human Research Ethics Committee of the Education University of Hong Kong. All data were obtained according to the principles expressed in the Declaration of Helsinki.

## 2.2. Stimuli

### 2.2.1. Speech stimuli

The speech stimuli (presented in Table 1) for the word learning trials consisted of 24 pairs of monosyllabic Cantonese (pseudo)words, differing by a minimal phonological contrast of 1 feature. Eight pairs involved a consonant contrast (e.g., /tœ6/ vs. /kœ6/), 8 involved a vowel contrast (e.g., /ku6/ - /kɔ6/), and 8 involved a tone contrast (e.g., /sœ1/ - /sœ6/). The tones of the target syllables in the consonant and vowel pairs varied across trials. For each contrast condition, the (pseudo)words had a CV structure in 4 pairs and a CVN structure in the other 4 pairs. All the stimuli were either (pseudo)words or low frequency words in Cantonese. The low frequency words (i.e., /t<sup>h</sup>im5/, /tim1/, /sim1/, /sy5/, /kyn6/) were mostly written words and none were listed in the Cantonese version of the MacArthur Communicative Development Inventory (Tardiff & Fetcher, 2008).

Table 1. Minimal pairs of pseudowords.

Condition	Pair	Feature/ Tone change
Consonant contrasts	/pi3/ - /ti3/	place
	/tœ6/ - /kœ6/	place

	/p <sup>h</sup> an5/ - /t <sup>h</sup> an5/	place
	/t <sup>h</sup> im5/ - /k <sup>h</sup> im5/	place
	/sɔ6/ - /fɔ6/	place
	/man3/ - /nan3/	place
	/tim1/ - /sim1/	manner
	/pɛ2/ - /fɛ2/	manner
Vowel contrasts	/k <sup>h</sup> im3/ - /k <sup>h</sup> em3/	height
	/sy5/ - /sœ5/	height
	/ku6/ - /kɔ6/	height
	/lem1/ - /lam1/	height
	/kyn6/ - /kun6/	backness
	/tœ3/ - /tɔ3/	backness
	/t <sup>h</sup> in2/ - /t <sup>h</sup> yn2/	roundedness
	/kɛ6/ - /kœ6/	roundedness
Tone contrasts	/k <sup>h</sup> im1/ - /k <sup>h</sup> im3/	T1—T3; register
	/sœ1/ - /sœ6/	T1--T6; register
	/tɛ3/ - /tɛ6/	T3—T6; register
	/p <sup>h</sup> an2/ - /p <sup>h</sup> an5/	T2—T5; register
	/man2/ - /man3/	T2—T3; contour
	/fa4/ - /fa5/	T4—T5; contour
	/fɔ4/ - /fɔ6/	T4--T6; contour
	/sim5/ - /sim6/	T5—T6; contour

Note: T1 (High Level 55), T2 (High Rising 25), T3 (Mid Level 33), T4 (Low Falling 21), T5 (Low Rising 23), T6 (Low Level 22). Tones differing in register have a similar contour, while tones differing in contours were matched in register as much as possible.

#### Words for the training trials

Two pairs of monosyllabic familiar words (/mau1/ ‘cat’ - /kœu5/ ‘dog’, and /pɔ1/ ‘ball’ - /fa1/ ‘flower’) were used in the two training trials to help toddlers understand the task.

The (pseudo)words were presented in sentences in Cantonese. For familiarization, they were embedded in a script, and appeared in six different sentences (see Table 2). In the test phase, one of the two words was designated twice, in two sentences (see details in “animated cartoons” section below). All speech stimuli were recorded in an infant-directed manner by a female native adult speaker of Hong Kong Cantonese.

Table 2. The Cantonese script used in the test.

	Cantonese	Gloss	English translation
<i><u>Learning phase 1</u></i>	嘩/咦! 呢度有個X! 係/a3/, 呢個係X! 咦? 你睇下我嘅X去咗邊? 噏, X 喺呢度。 望下喺邊度/a3/ 我嘅X? 喺呢度/wɔ3/, 個X!	wow/huh, here has a X yes /sfp/*, this is X huh, you see my X gone where see, X is here see (it) is where /sfp/ my X (it) is here /sfp/, the X	Wow/Huh! Here is a X! Yes, this is X! Huh? Where do you see my X has gone? See, X is here. Do you see where my X is? The X is here!
<i><u>Learning phase 2</u></i>	咦/嘩! 呢度有個Y! 係/a3/, 呢個係Y! 咦? 你睇下我嘅Y去咗邊? 噏, Y 喺呢度。 望下喺邊度/a3/ 我嘅Y? 喺呢度/wɔ3/, 個Y!	huh/wow, here is a Y yes/sfp/, this is Y huh, you see my Y gone where see, Y is here see (it) is where /sfp/ my Y (it) is here /sfp/, the Y	Huh/Wow! Here is a Y! Yes, this is Y! Huh? Where do you see my Y has gone? See, Y is here. Do you see where my X is? The X is here!

<i>Close-up</i>	嘩/咦, 你(再)睇下!	wow/huh, you (again) see see	Wow/Huh, you see (again)!
<i>Test phase</i>	睇下個X/Y! X/Y喺邊度/a3/?	look (at) the X/Y X/Y is where /sfp/	Look at the X/Y! Where is the X/Y?

\*/sfp/: sentence-final-particle

### 2.2.2. Object Stimuli

Images of eight pairs of objects differing in shape and color (see Figure 2) were adapted from Gonzalez-Gomez, Poltrock and Nazzi (2013). The reason for using clearly different objects was to facilitate learning of the word-object pairings. All objects were selected so that they would look novel to the participants. The same 8 object pairs were used in the 3 contrast conditions (consonant, vowel, tone). This was done in order to ensure that overall performance differences across conditions could not be due to the objects used.

Images for the training trials

We also used 4 color images to depict the 4 familiar words (cat, dog, ball and flower) in the training trials (see Figure 2).

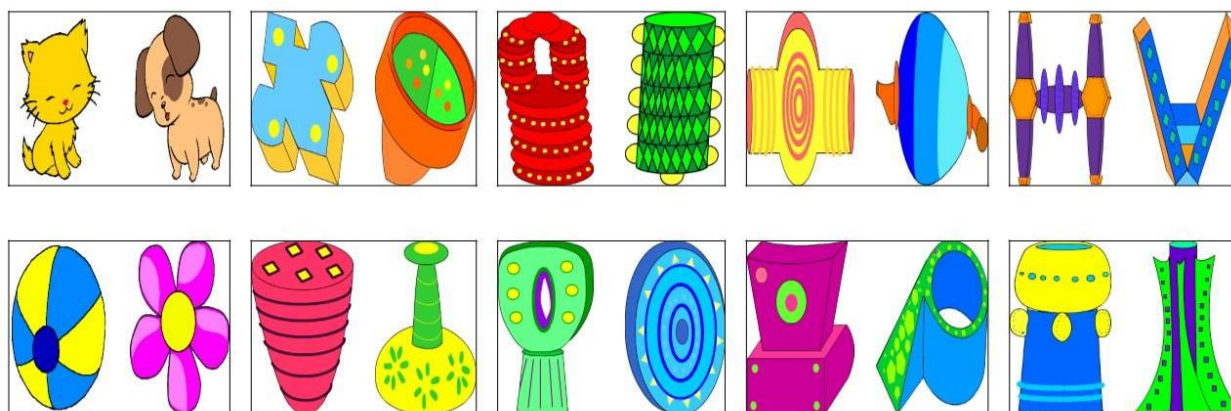


Figure 2. The 2 pairs of familiar objects (cat-dog, ball-flower) used for training and the 8 pairs of novel objects used for word learning.

### *Animated cartoons*

The audio recordings were included in animated cartoons slightly adapted from cartoons that have been successfully used in a computer-controlled word-learning task in toddlers by Gonzalez-Gomez et al. (2013), and in adults by Poltrock et al. (2018). In the current version, the animations of the human character were reduced so as not to distract toddlers' attention to the target objects, while the animations of the target objects were increased. An example of a cartoon is illustrated in Figure 3.

On each trial, a female cartoon experimenter behind a blackboard presented the two objects, one at a time (Fig. 3, learning phase). The first object always appeared first in the left upper corner of the cartoon. At the beginning, the object moved horizontally in the upper left part of the cartoon and spun, while it was labeled twice ("Wow! Here is a [label]! Yes, this is [label]!"). Then, the object started enlarging and shifting towards the center of the left part of the cartoon, while the blackboard was shifting up and taking up the whole background. The object was labeled again ("Huh? Where do you see my X [label] has gone? ") at the same time. It then started spinning and moving vertically in the left part of the movie frame and was labeled three more times ("See, [label] is here. Do you see where the [label] is? The [label] is here!") before disappearing. The second object was always introduced in the upper right corner of the movie at the beginning and followed a trajectory analogous to that of the first object. The cartoon experimenter first looked at the front and then clearly moved her eye gaze towards the object on the left/right after it appeared and before it started moving. Participants were successively trained on each label-object pairing for 25 seconds. The entire learning phase lasted 50 seconds, during which each label was repeated 6 times.

After the learning phase, participants were tested immediately on the given contrast. There was a close-up on the face of the cartoon experimenter saying: "Wow/Huh, you see!" in order to direct the participants' fixations to the center of the screen. After the face had disappeared, the two objects appeared at the same time, each on the side where it had appeared during the learning phase, and started moving synchronously in a vertical way, for 6367 ms, while the out-of-sight speaker said: "Look at the [target]? Where's the [target]!" in which the



first mention of the target word appeared about half way through the presentation, at 3000 ms (Fig. 3, test phase). Since it has been shown that it takes 367 ms for infants and toddlers to program eye movements after hearing a target word (e.g., Swingley & Aslin, 2000), and since this delay has been used in many prior infant studies on word recognition/learning (e.g., Swingley & Aslin, 2002, 2002; Mani & Plunkett, 2007; White & Aslin, 2011; Gonzalez Gomez et al., 2013), the *3s postnaming phase* started at 3367ms for all conditions (hence 3367-6367ms), and was immediately preceded by a *3s prenaming phase* (367-3367ms). Note that using such a delay should, if anything, facilitate recognition in the consonant trials, for which the contrasting information would become available sooner).

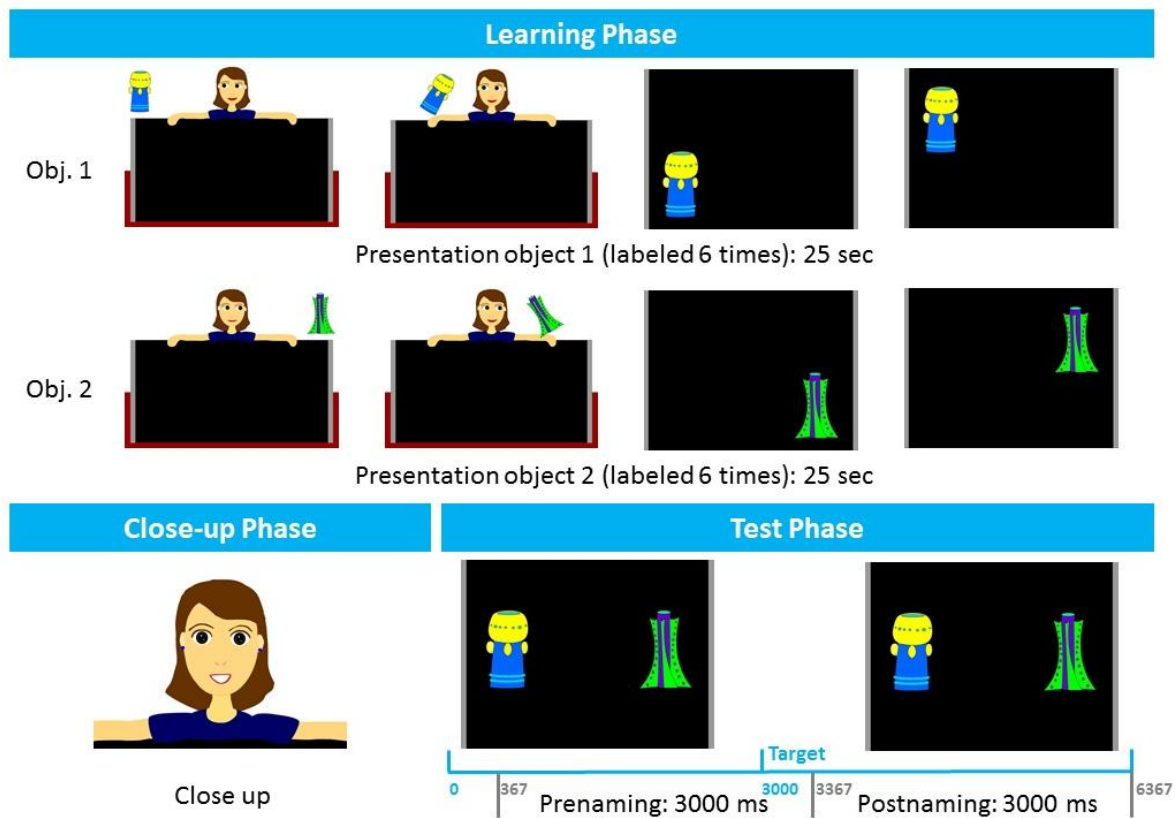


Figure 3. Structure of a word-learning cartoon.

Every object pair was associated with one pseudoword pair in each testing condition (e.g., object A and B were associated with /pi3/ and /ti3/ in the consonant condition; with /k<sup>h</sup>im3/ and /k<sup>h</sup>em3/ in the vowel condition and with /k<sup>h</sup>im1/ and /k<sup>h</sup>im3/ in the tone condition), for use in 24 different trials (8 trials x 3 conditions). Eight versions of each cartoon were created so that for

each word pair, each of the two labels was paired with each of the two objects 50% of the time. In half of the trials, object/label A was the target (and consequently object/label B the distractor) and, in addition, the target was presented as first object in 50% of the trials and as second object in the other 50%. This yielded a total stimulus set of 192 movies, all having a resolution of 1024 x 768 pixel. Presentation of each of the eight versions of each cartoon was counterbalanced across participants.

Similar animated cartoons were also made and used for the 2 training trials.

### **2.3. Apparatus and Procedure**

A Tobii TX300 was used to present the cartoon movies at the center on a 23'' integrated screen unit with a 1920 x 1080 pixel resolution. The eye-tracker was tilted at 25°. The frame of the cartoon movies took up a size of 27.2 x 20.5 cm. An integrated camcorder above the screen monitored the participants' behavior. The presentation of the stimuli and the storing of the data were performed with the Tobii Studio software (version 3.4.8).

Participants were tested individually in a dimly lit, quiet laboratory room. Each toddler sat approximately 60-65 cm from the screen on a caregiver's lap in the center of the test booth. The caregiver was wearing opaque glasses to prevent them from seeing the stimuli and thus minimize the potential for biases. The experimenter controlled the presentation of the stimuli from an adjacent area and monitored the participant's behavior through a video camera. The session began with a 5-point infant calibration. Then a small animation was displayed on the center of the screen before each of the 8 trials until the child looked at it, in order to start each trial at the center of the screen.

Each of the 8 test trials had the same structure, and corresponded to the presentation of a cartoon. Each trial was thus composed of the learning of 2 (pseudo)words, immediately followed by a close-up and a testing phase evaluating learning/recognition. In the test phase of each trial, infants were required to look at one of the two objects. Therefore, within each trial, one object was the target and the other one was the distractor.

There were eight pseudo-randomized orders counterbalancing for target side, target object, trial order and object label. Thus, between-subjects counterbalancing ensured that each label was presented and tested on the left and right side. Within-subjects, the first and the second half of the test trials always contained 2 CV and 2 CVN trials, and half of the time the target

word was on the left, half of the time it was on the right. No more than 3 consecutive trials would share the same phonological contrast in feature, or have targets on the same side. The experiment lasted approximately 10 minutes.

#### **2.4. Data Analysis and Exclusion Criteria**

The eye-tracking data used for the analysis consisted of the binocular gaze position (X and Y coordinates) at each timestamp, that is, every 3.3 msec. First, the proportion of on-screen looks during the course of the 8 test trials was calculated for each trial. To ensure that infants were sufficiently engaged in the task, we excluded 187 trials (27.5% of total) with less than 50% on-screen data during the test phase at 20 months, and 176 trials (27.1% of total) infants with less than 50% on-screen data during the test phase at 30 months.

For each kept trial, we then calculated the proportion of time infants spent looking at the target (T) and the distractor (D) in both the pre-naming and the post-naming phases. Therefore, two areas of interest were defined (460 x 768 Pixel), each including one object. We excluded 12 additional trials (1.8% of total) in which infants had a strong object bias in the pre-naming phase (90% looking at one object) at 20 months, and 26 additional trials (4.0% of total) in which infants had a strong object bias in the pre-naming phase at 30 months. This was done in order to remove trials in which the baseline preference was not well established, as done by Gonzalez Gomez et al. (2013).

Finally, only those infants who had at least four analyzable trials were included to have a better estimate of each infant's "true" naming effect (13 20-month-olds and 9 30-month-olds did not meet this criterion; for a similar criterion, see also Gonzalez Gomez et al., 2013). In the final sample, each participant provided, on average, 6.49 trials out of 8 at 20 months, and 6.03 trials out of 8 at 30 months.

#### **2.5. Vocabulary Measure**

To determine the size of the infants' receptive vocabulary, parents were asked to fill out the vocabulary part of the Cantonese equivalent (Tardiff & Fletcher, 2008) of the MacArthur Communicative Development Inventory: Toddlers (Fenson et al. 1991).

### 3. Results

#### 3.1. Accuracy- Overall analysis

We first calculated the mean proportion of target looking (PTL = total looking time to target/ total looking time to both objects) on each trial for both the pre- and postnaming phases. Time stamps that were not in any of the AOIs were treated as missing data, so that the calculated proportion of looking to one AOI is always relative to both AOIs, resulting in values between 0 and 1 (i.e., a proportion value of 0.5 means that each AOI was looked at equally long). PTLs were then averaged for each toddler, and then for each condition/type of contrasts at each age (see Figure 4).

A repeated-measures ANOVA on PTL with *naming* (pre- vs. postnaming phase) as within-subject factor, *condition* (consonant- vs. vowel- vs. tone-contrasted) and *age* (20- vs. 30-month-olds) as between-subject factors was conducted. There was a significant effect of *naming* ( $F(1, 138) = 6.026, p = .015, \eta p^2 = .042$ ) and a significant *naming* x *age* interaction ( $F(1, 138) = 4.014, p = .047, \eta p^2 = .028$ ); all other effects failed to reach significance. Follow-up ANOVAs were conducted at each age, with *naming* as within-subject factor and *condition* as between-subject factor. At 20 months, no effects reached significance. At 30 months, there was a significant *naming* effect ( $F(1, 69) = 9.581; p = .003, \eta p^2 = .122$ ). Because we were interested in comparing learning effects in the different conditions, and although the *naming* x *condition* interaction failed to reach significance ( $F(2, 69) = .882; p = .419, \eta p^2 = .025$ ), the naming effect was analyzed separately for the 3 conditions: it failed to reach significance for both the C ( $t(23) = .955, p = .35, 2$ -tailed) and T ( $t(23) = 1.692, p = .10, 2$ -tailed) conditions, but was significant for the V condition ( $t(23) = 2.786, p = .01, 2$ -tailed), even when controlling for multiple comparisons ( $\alpha = .05/3 = .0167$ ).

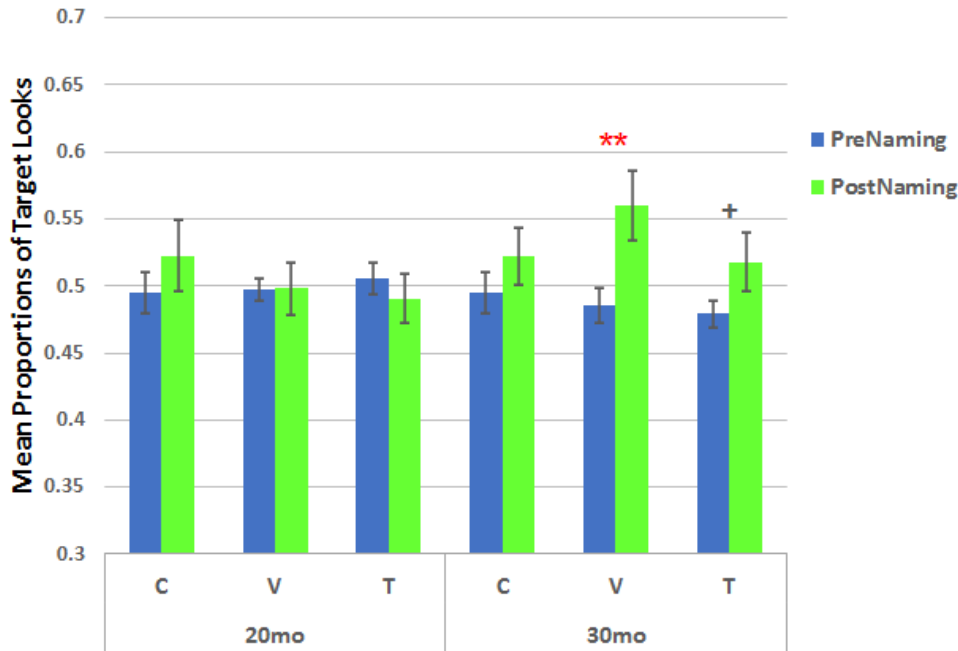


Figure 4. Results for the PTL analysis (Means and SEs), broken down by age and condition. \*\*  $p < .01$ ; +  $p = .052$ .

### 3.2. Correlation between PTL difference and CDI

Children produced a mean of 201 words ( $SD = 163$ ) at 20 months, and a mean of 598 words ( $SD = 157$ ) at 30 months, a significant difference,  $t(140) = 14.764$ ,  $p < .001$ , 2-tailed. There was no difference between the 3 conditions at 20 months ( $F(2, 67) = 1.733$ ,  $p = .185$ ) and 30 months ( $F(2, 69) = .601$ ,  $p = .551$ ). To determine whether learning performance is linked to the number of words produced, we performed Spearman's Correlations between the recognition effect ( $PTL_{postnaming} - PTL_{prenaming}$ ) and the number of words at test, for each of the 6 groups separately. All correlations failed to reach significance (see Table 3).

Table 3. Results of the correlations

	20-month-olds			30-month-olds		
	r	n	p (2-tailed)	r	n	p (2-tailed)
Consonant contrasts	-0.144	23	.513	0.277	24	.189

Vowel contrasts	-0.014	23	.950	-0.397	24	.055
Tone contrasts	0.333	24	.112	0.257	24	.226

The above results establish that while 30-month-olds succeeded in the task, 20-month-olds did not, suggesting the task might be too difficult for them; and at both ages, we found no evidence that performance correlated with vocabulary size. Given the overall failure of the 20-month-olds, we did not further analyze the data of this age group. At 30 months, the PTL results suggest some differences in performance between the different conditions, with best performance for the vowel-contrasted condition, worst performance for the consonant-contrasted condition, and intermediate performance for the tone-contrasted condition. Yet, averaging performance over the entire pre- and postnaming phases might have obscured the differences, leading to the nonsignificant interaction. In what follows, we further evaluate differences in performance across conditions, using growth curve analyses of the time course of toddlers' response to naming.

### 3.3. Time course Analysis (30-month-olds only) - Growth Curve Analysis

Figure 5 presents the raw averaged target looking data for the entire test phase for the 30-month-olds, broken down by condition. At that age, we conducted a Growth Curve Analysis (GCA) which includes time as a predictor to estimate if differences between conditions emerged over time within each language group. As dependent measure, we took the transformed proportion data during the postnaming phase using the empirical logit (*elog*, aggregated in 100 ms time bins) and analyzed it with a weighted mixed-effects linear regression model within the *eyetrackingR* package (modeled after Mirman, Dixon, & Magnuson, 2008). We entered condition (compared in sliding contrasts: Consonant-Vowel; Vowel-Tone), orthogonal polynomials (linear, quadratic and cubic time component), and the interaction between each time term and condition as fixed effects. The tested (pseudo)word pairs and participants were entered as random effects into the model. The formula used was:

$$\text{Elog} \sim \text{Condition} * (\text{ot1} + \text{ot2} + \text{ot3}) + (1 \mid \text{WordPair}) + (1 \mid \text{Participant})$$

(ot1, ot2, and ot3 represent the linear, quadratic, and cubic time term, respectively)

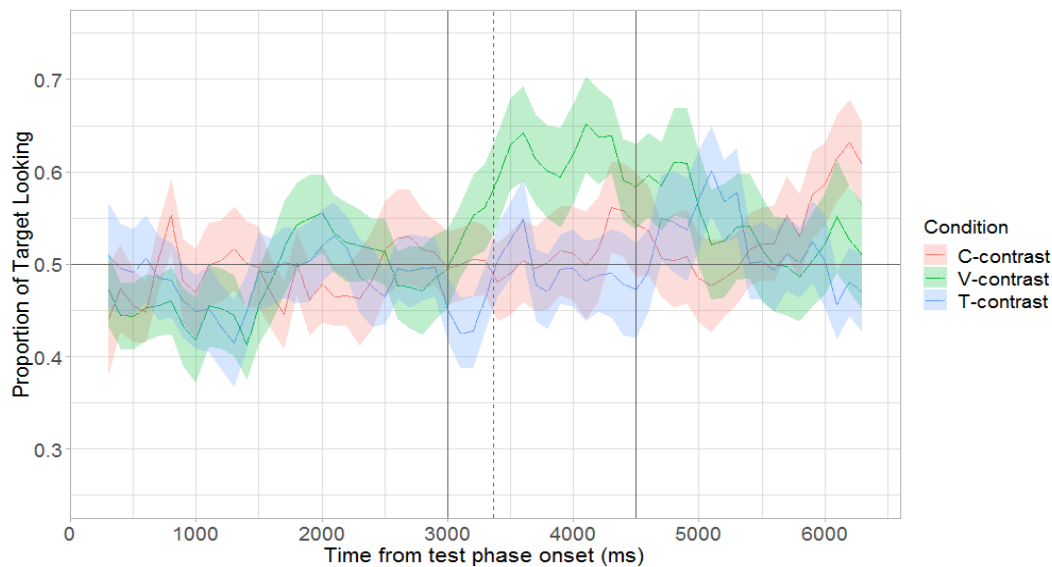


Figure 5. Target looking behavior during the test phase in the three conditions, for the 30-month-olds only. The solid lines refer to the onset of the target word at the first mention (time point 3000) and at the second mention (time point 4500). The dashed line represents the beginning of the postnaming phase.

The results of the model are presented in Figure 6, and the statistical results for the fixed effects in Table 4. There was no significant main effect of the Consonant-Vowel ( $\beta = 0.317$ ,  $SE = 0.329$ ,  $p = .337$ ) and Vowel-Tone ( $\beta = -0.346$ ,  $SE = 0.329$ ,  $p = .293$ ) contrasts, indicating no differences in overall target looking in the postnaming phase between those conditions, confirming the PTL results. However, we found a significant interaction between the Consonant-Vowel contrast and time (specifically, the linear parameter:  $\beta = -2.428$ ,  $SE = .468$ ,  $p < .001$ ), and between the Vowel-Tone contrast and time (linear parameter:  $\beta = 1.608$ ,  $SE = .467$ ,  $p = .001$ ; cubic parameter:  $\beta = -1.886$ ,  $SE = .465$ ,  $p < .001$ ), suggesting divergent linear temporal trajectories for vowel and consonant trials, as well as divergent linear and nonlinear temporal trajectories for tone and vowel trials in the postnaming phase. For the Consonant-Vowel contrast, this suggests that the difference in response is larger at the beginning and decreases with time (linear effect). For the Vowel-Tone contrast, this suggests that the difference in response also decreases with time (linear effect). The interpretation of the cubic effect is more difficult to interpret, but suggests that variation in the direction of looking at the target versus distractor changes differently (3 changes in focus) for the two conditions as the postnaming phase unfolds.

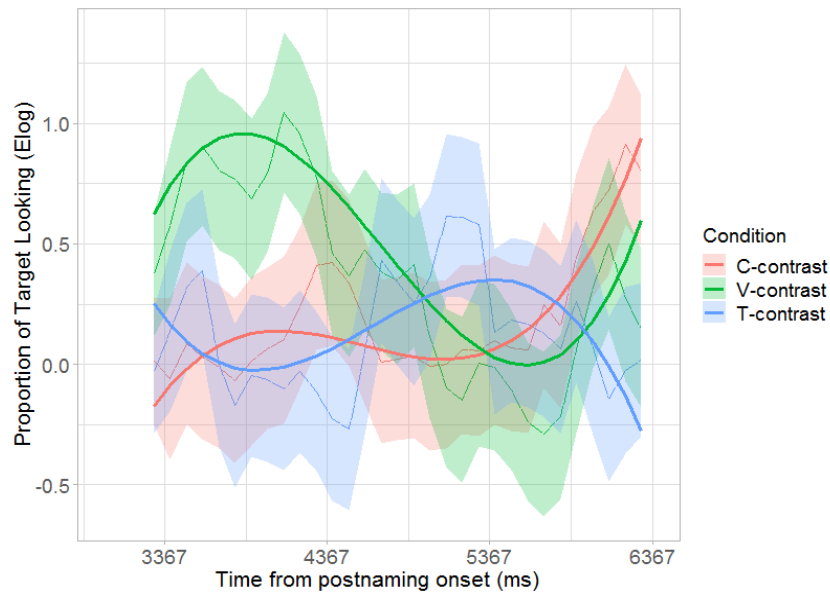


Figure 6. Time course during the postnaming phase of consonant, vowel and tone trials for the 30-month-olds; shown as raw data (light) and fitted curves (bold).

Table 4. Statistical output of the fixed effects in the growth curve analysis for the postnaming phase of the 30-month-olds.

	Estimate	Std. Error	t value	p
(intercept)	0.260	0.134	1.936	0.053
conditionC-V	0.317	0.329	0.961	0.337
conditionV-T	-0.346	0.329	-1.052	0.293
ot1	-0.152	0.191	-0.798	0.425
ot2	0.184	0.191	0.961	0.337
ot3	0.357	0.191	1.874	0.061
conditionC-V : ot1	-2.428	0.468	-5.193	0.000



conditionV-T : ot1	1.608	0.466	3.448	0.001
conditionC-V : ot2	-0.242	0.468	-0.517	0.605
conditionV-T : ot2	-0.761	0.466	-1.631	0.103
conditionC-V : ot3	0.415	0.468	0.885	0.376
conditionV-T : ot3	-1.886	0.465	-4.059	0.000

---

#### 4. Discussion

In this study, we investigated whether 20- and 30-month-old children can quickly learn new minimal pair words in their native language, Cantonese, a tone language. To specify how, in this tone language, native phonological knowledge interacts with word learning, the Cantonese pseudowords within a pair differed minimally in either a consonant, a vowel, or a tone. Overall, we found that 20-month-olds could not learn the words, but 30-month-olds did. However, in these older children, learning differed across the different types of contrasts. When analyzing increases in looking times between the pre- and post-naming phases (PTL analysis), we found a significant increase only for the vowel condition. When taking into account processing time, we found on the postnaming phase better performance in the vowel condition compared to both the consonant and tone conditions. This was found even though the fact that the postnaming phase started at the onset of the target word should have given an advantage to the consonant condition, as the contrasting information in this condition arrives earlier than in the other two conditions. Lastly, performance did not correlate with vocabulary levels in none of the 6 age/condition groups.

The present findings first establish that, using this procedure, learning words was only possible at 30 months, not 20 months. Since many studies have found that it is possible to learn words in the laboratory, including at younger ages (e.g., Stager & Werker, 1997; Havy & Nazzi, 2009), the present failure at 20 months has to be linked to task issues. Several factors could have made task difficulty too high in our paradigm. First, the use of 1-feature contrasts required children to process fine phonetic information while pairing the objects with the labels. Second,

because we wanted to test children on 8 different contrasts, they were presented with only 6 repetitions of the object-label associations to reduce the length of the experiment. Since learning in such conditions has been found in many experiments, even by 16 months (e.g., Havy & Nazzi, 2009), these two factors alone cannot fully explain the present 20-month-olds' failure. A third factor is that learning and testing was fully conducted through non-infant-controlled screen presentation (the trials were launched when the child looked at the screen but continued to the end independently of their behavior) rather than live interaction, although such conditions have been found to be a more difficult learning condition in some studies (e.g., Kuhl, Tsao & Liu, 2003). The combinations of these three experimental factors might have made learning too hard for 20-month-olds. To further test this age group of 20-month-olds, simplifications of the task will be needed in the future.

Focusing now on the 30-month-olds, our finding of a vowel advantage, compared to both consonants and tones, bears new evidence regarding two issues: the relative weight of consonants and vowels for lexical processing in a tone language, and the comparison of the processing of tonal versus segmental (consonant and vowel) information.

Starting with the consonant/vowel comparison, our findings provide clear evidence of a vowel bias at 30 months in Cantonese. The combination of the PTL and growth curve analyses indicates that infants only learn in the vowel condition, and that word learning performance is better in the vowel than in the consonant condition when taking the dynamics of eye gaze movements in the postnaming phase into account. The differences between the two conditions appear particularly clear in the first half of the postnaming, a finding that also demonstrates rapid recognition of the named target, and thus of the processing of vocalic information. An inspection of the time-course analysis even suggests that recognition might actually start earlier than 367ms after word onset, as suggested by Swingley and Aslin (2000) for 18-to-23-month-olds. This would be compatible with findings suggesting an increase in word recognition speed across the second year of age (Fernald, Perfors & Marchman, 2006). This finding of a V-bias in Cantonese is opposite to what has been found in most non-tone languages (to the exception of Danish-learning 20-month-olds, Højen & Nazzi, 2016): by 30 months, C-biases in lexical processing have been acquired in French, English and Italian (Nazzi et al., 2009; Hochmann et al., 2011). In the introduction, we had discussed that the presence of lexical tones might affect the relative weight of consonants and vowels in either one direction or the other. It could have strengthened

the C-bias, as vowels are likely to be acoustically more variable in tone languages, carrying the tones. In contrast, our preferred hypothesis was that it could have reversed the bias, by drawing attention to vowels that would carry both quality (segmental) and tonal (suprasegmental) information. Our developmental findings of a clear V-bias in Cantonese, in a population of 30-month-old children, support the latter prediction, and clearly establish that, in this tone language, the interplay between phonological and lexical acquisition follows a different developmental trajectory from that of most European languages.

These findings are in line with evidence from adult studies on phonological biases in Cantonese, in which Cantonese-speaking adults failed to show a C-bias in a word segmentation study using an artificial language (Gomez et al., 2017), or a new word learning experiment (Poltock et al., 2018), suggesting a different balance in the relative weight of consonant and vowel processing in this language, to the advantage of vowels (although no clear V-bias has yet been found in Cantonese-speaking adults). Our findings are also in line with both adult and developmental studies in another tone language, Mandarin. Recent studies suggest that Mandarin adults have a V-bias in word reconstruction tasks, when processing either Mandarin or English stimuli (Wiener & Turnbull, 2016; Wiener, 2020), and fail to show a C- or V-bias in new word learning tasks (Poltock et al., 2018). Developmental studies on Mandarin found that while no difference in Mandarin-learning and bilingual Mandarin-English 6-year-olds (Wewalaarachchi & Singh, 2020), bilingual Mandarin-English toddlers (2.5–3.5 years) and preschoolers (4–5 years; Singh et al., 2015), or Mandarin-learning and bilingual Mandarin-English 24-month-olds when analyzing PTL, a V-advantage was found when analyzing only distractor-initial trials at that age (Wewalaarachchi et al., 2017). Taken together, the studies on Mandarin suggest increased weight to vowels compared to consonants in this other tone language, although clear V-biases appear difficult to establish. Note also that because the effects appear to change over development in Mandarin (but also in French, Italian and English, see Introduction), follow-ups to the present study should aim at tracing the developmental trajectory of the bias in Cantonese, as was our original plan by testing 20- and 30- month-olds.

Why did we find a clear V-bias in tone language Cantonese? Previous studies have suggested that the emergence of phonological biases at the lexical level is influenced both by phonological and lexical properties of the native language (see Nazzi & Cutler, 2019, for a review). At present, it remains unclear what exact factors induce more weight towards vowels in

Cantonese. Lexical factors such as the ratio of words that can be contrasted by a C or a V in that language would need to be explored. In terms of phonological factors, the presence of 6 different tones, which are (mostly) carried by the vowels, might first multiply the number of (phonetic) vowels, so that this C/V ratio might tilt towards more vowels, supporting a V-bias. Second, the variation brought by the tones might also have an attentional effect, bringing learners and speakers of Cantonese to pay more attention to vowels due to their increased variability. These two effects would suggest that the increased variability of the acoustic realization of vowels due to the presence of tones would help, rather than hinder, vowel processing. Third, if vowels have to be considered together with their tonal information (at least for some processing aspects), this might also have an effect at the lexical level, since two words with the same CV content but differing in their tone would be considered as different in terms of vocalic content, which would support a V-bias. Note that the fact that we might have found a clearer V-bias than could be found in Mandarin (although a direct comparison is not possible, as similar experiments were not conducted in both languages: new word learning in our Cantonese study; sensitivity to mispronunciations of familiar words in in the Mandarin studies), which has a simpler tone system (4 instead of 6 tones), if confirmed, would also support the role of tones in determining V-biases. Phonological factors other than tones might also play a role, such as the facts that Cantonese only allows simple syllable structures (e.g., no consonant clusters), that only a few of its consonants (/j, w, m, n, ŋ/) are voiced, and that little vowel reduction occurs, will have to be further investigated.

We now discuss the comparison of the processing of tonal and segmental information while learning new words at 30 months. Both the PTL and the growth-curve analyses revealed that the tone-contrasted condition was harder than in the vowel-contrasted condition. This suggests that although vocalic and tonal information are both carried by the vowel, they are processed differently and possibly independently. Moreover, our findings appear to differ from the pattern found for Mandarin, in which differences between the processing of tonal and segmental information were always found. In that language, least sensitivity to tone mispronunciations was found in older children: Mandarin-learning and bilingual Mandarin-English 6-year-olds, and Mandarin-learning 4-to-5-year-olds, and bilingual Mandarin-English 24-month-olds in the time course analysis (Singh et al., 2015; Wewalaarachchi et al., 2017; Wewalaarachchi & Singh, 2020). On the contrary, most sensitivity to tone over both consonant

and vowel mispronunciations was found in Mandarin-learning 2.5-to-3.5-year-olds and 24-month-olds in the time course analysis (Singh et al., 2015; Wewalaarachchi et al., 2017). Given the complexity of the pattern found in Mandarin, and the fact that we only have data for one group of monolingual 30-month-olds, specifying what the differences are between the two tone languages proves premature. Future studies will have to provide parallel data, using the same methods, of comparative processing of tonal versus segmental information at different ages in order to start understanding the factors that might modulate this comparison, whether they are due to tonal differences between the two languages, or other phonological and lexical factors.

In conclusion, our findings from a tone language, Cantonese, showing that phonological biases in lexical processing vary across languages and that a C-bias is not always found (it has also not been found in 20-month-old Danish-learning infants, Højen & Nazzi, 2016) have consequences for our theoretical understanding of the link between phonological and lexical acquisition. The studies in this domain received a crucial impulse from the proposal of the division of labor hypothesis (Nespor et al., 2003), according to which there would be a C-bias in lexical processing, and a V-bias in syntax-processing. Do our findings disprove the division of labor hypothesis? We do not think this is the case, as illustrated by a recent review showing that of 38 adult studies investigating this issue in 15 languages (including 3 varieties of English), most find a C-bias, with ambiguous findings only for Berber, Mandarin and Cantonese (Nazzi & Cutler, 2019). Hence, a C-bias in lexical processing seems to be the predominant pattern found in adulthood. Yet, that same review further highlighted variability in the outcome of 27 developmental studies investigating this issue in 7 languages (French, Italian, Danish, Mandarin, and 3 varieties of English), with some languages not showing a clear and early C-bias, and also a tendency for initial V-biases to change into C-biases with native language acquisition. Our results contribute to this line of work, thus supporting the acquired bias hypotheses over the innate bias hypothesis. Clearly, more cross-linguistic studies need to be conducted, in order to differentiate the 3 different acquired bias hypothesis (acoustic/phonetic hypothesis, Floccia et al., 2014; lexical hypothesis, Keidel et al., 2007; phono-lexical hypothesis, Nishibayashi & Nazzi, 2016; Poltrock & Nazzi, 2015), and trace the developmental trajectory of these biases across the lifespan, in order to determine whether the early V-biases found in Danish (Højen & Nazzi, 2016), Mandarin (Wewalaarachchi et al., 2017) and Cantonese (present study) persist later in life. Importantly also, what the Nespor et al. (2003) proposal has brought to our attention is that the

difference between consonants and vowels that has been noted for centuries has functional implications, and that these have a strong impact on language processing from infancy to adulthood. It led to 20 years of research that have started to specify the acquisition of the phonological-lexical and phonological-syntactic interfaces, and is likely to continue to inspire further studies in this domain.

## REFERENCES

- Abboub, N., Boll-Avetisyan, N., Bhatara, A., Höhle, B. & Nazzi, T. (2016). An exploration of rhythmic grouping of speech sequences by French- and German-learning infants. *Frontiers in Human Neuroscience*, 10:292. DOI: 10.3389/fnhum.2016.00292
- Bauer, R.S., & Benedict, P.K. (1997). *Modern Cantonese Phonology*. Berlin and New York: Mouton de Gruyter. doi.org/10.1515/9783110823707
- Benavides-Varela, S., Hochmann, J.R., Macagno, F., Nespors, M., & Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences*, 109(44), 17908-17913. 10.1073/pnas.1205413109
- Bergelson, E., & Swingle, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9), 3253–3258. DOI:10.1073/pnas.1113380109
- Bonatti, L.L., Peña, M., Nespors, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychological Science*, 16, 451–9. 10.1111/j.0956-7976.2005.01556.x
- Bouchon, C., Floccia, C., Fux, T., Adda- Decker, M., & Nazzi, T. (2015). Call me Alix, not Elix: vowels are more important than consonants in own- name recognition at 5 months. *Developmental science*, 18(4), 587-598. doi.org/10.1111/desc.12242
- Cheng, R.L. (1966). Mandarin phonological structure. *Journal of Linguistics*, 2, 135-158. doi.org/10.1017/S0022226700001444
- Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: cross-linguistic comparisons. *Memory & Cognition*, 28, 746–55. doi.org/10.1177/0023830915578000
- DeCasper, A.J., & Fifer, W.P. (1980). Of human bonding: Newborns prefer their mother's voices. *Science*, 208, 1174-1176. DOI: 10.1126/science.7375928
- DeCasper, A.J. & Spence, M.J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, 9(2), 133-150. doi.org/10.1016/0163-6383(86)90025-1
- DePaolis, R.A., Vihman, M.M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, 34, 590-601. doi.org/10.1016/j.infbeh.2011.06.005

- Duanmu, S. (2000). *The Phonology of Standard Chinese*. Oxford: Oxford University Press. ISBN: 9780199215799
- Eimas, P.D., Siqueland, E.R., Jusczyk, P.W., & Vigorrito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306. doi: 10.1126/science.171.3968.303
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., ... Reilly, J. S. (1991). *Technical manual for the MacArthur communicative development inventories*. San Diego, CA: San Diego State University. ISBN 978-1-55766-884-4
- Fernald, A., Perfors, A. & Marchman, V.A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98-116. DOI: 10.1037/0012-1649.42.1.98
- Floccia, C., Nazzi, T., Delle Luche, C., Poltrock, S., & Goslin, J. (2014). English-learning one-to-two-year-olds do not show a consonant bias in word learning. *Journal of Child Language*, *41*(5), 1085-1114. DOI: 10.1017/S0305000913000287
- Gómez, D.M., Mok, P., Ordin, M., Mehler, J., & Nespors, M. (2017). Statistical speech segmentation in tone languages: The role of lexical tones. *Language and Speech*, *61*(1), 84–96. DOI: 10.1177/0023830917706529
- Gonzalez-Gomez, N., Poltrock, S., & Nazzi, T. (2013). A “bat” is easier to learn than a “tab”: Effects of relative phonotactic frequency on infant word learning. *PlosOne* *8*(3), e59601. DOI: 10.1371/journal.pone.0059601
- Hashimoto, A.O.-K.Y. (1972). *Phonology of Cantonese*. Cambridge: Cambridge University Press. ISBN 0521084423 9780521084420
- Havy, M., Serres, J., & Nazzi, T. (2014). A consonant/vowel asymmetry in word-form processing: Evidence in childhood and in adulthood. *Language and Speech*, *57*, 254–281. doi.org/10.1177/0023830913507693
- Havy, M. & Nazzi, T. (2009). Bias for consonantal over vocalic information in word learning at 16 months. *Infancy*, *14*, 439-456. DOI: 10.1080/15250000902996532
- Hochmann, J.R., Benavides- Varela, S., Nespors, M., & Mehler, J. (2011). Consonants and vowels: different roles in early language acquisition. *Developmental science*, *14*(6), 1445-1458. doi.org/10.1111/j.1467-7687.2011.01089.x
- Hochmann, J.R., Benavides- Varela, S., Fló, A., Nespors, M., & Mehler, J. (2018). Bias for vocalic over consonantal information in 6- month- olds. *Infancy*, *23*(1), 136-151. doi.org/10.1111/infa.12203
- Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J. & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development*, *32*, 262-274. DOI: 10.1016/j.infbeh.2009.03.004
- Højten A., & Nazzi T. (2016). Vowel bias in Danish word-learning: Processing biases are language-specific. *Developmental Science*, *19*, 41–49. doi.org/10.1111/desc.12286
- Howie, J. (1976). *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press. ISBN: 9780521134255

- Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development, 64*, 675–687 DOI: 10.2307/1131210
- Keidel, J.L., Jenison, R.L., Kluender, K.R. & Seidenberg, M.S. (2007). Does grammar constrain statistical learning? Commentary on Bonatti, Peña, Nespor and Mehler (2005). *Psychological Science 18*, 922–23. doi.org/10.1111/j.1467-9280.2007.02001.x
- Kuhl, P.K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America, 100(15)*, 9096-9101. doi:10.1073/pnas.1532872100
- Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606-608. DOI: 10.1126/science.1736364
- Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S., Dubois, J., Grebe, R., & Wallois, F. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proceedings of the National Academy of Sciences, 110(12)*, 4846-4851. doi/10.1073/pnas.1212220110
- Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language, 57*, 252–272. doi.org/10.1016/j.jml.2007.03.005
- Mani, N., & Plunkett, K. (2010). Twelve-month-olds know their cups from their keps and tups. *Infancy, 15*, 445–470. doi.org/10.1111/j.1532-7078.2009.00027.x
- Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*, 241-265. doi.org/10.1207/s15327078in1003\_3
- Mehler, J., Bertoncini, J., Barriere, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception, 7*, 491-497. DOI: 10.1068/p070491
- Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*, 143-178. doi.org/10.1016/0010-0277(88)90035-2
- Mirman, D., Dixon, J.A., & Magnuson, J.S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59(4)*, 475-494. doi: 10.1016/j.jml.2007.11.006. doi.org/10.1016/j.jml.2007.11.006
- Moon, C., Panneton-Cooper, R., & Fifer, W.E. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*, 495-500. doi.org/10.1016/0163-6383(93)80007-U
- Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition, 98(1)*, 13-30. doi.org/10.1016/j.cognition.2004.10.005



- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 756-766. DOI: 10.1037//0096-1523.24.3.756
- Nazzi, T., & Cutler, A. (2019). How consonants and vowels shape spoken-language recognition. *Annual Review of Linguistics*, *5*, 25–47. doi.org/10.1146/annurev-linguistics-011718-011919
- Nazzi, T., Poltrock, S., & Von Holzen, K. (2016). The developmental origins of the consonant bias in lexical processing. *Current Directions in Psychological Science*, *25*(4), 291-296. doi.org/10.1177/0963721416655786
- Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, *21*, 779-784. doi.org/10.1016/S0163-6383(98)90044-3
- Nazzi T., Floccia C., Moquet B., & Butler J. (2009). Bias for consonantal information over vocalic information in 30-month-olds: Cross-linguistic evidence from French and English. *Journal of Experimental Child Psychology*, *102*, 522–537. doi.org/10.1016/j.jecp.2008.05.003
- Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue E Linguaggio*, 203–230. DOI: 10.1418/10879
- Newman, R., Ratner, N.B., Jusczyk, A.M., Jusczyk, P.W., Dow, K.A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: a retrospective analysis. *Developmental Psychology*, *42*, 643-655. DOI: 10.1037/0012-1649.42.4.643
- Nishibayashi, L.-L. & Nazzi, T. (2016). Vowels then consonants: emergence of a consonant bias in early word form segmentation. *Cognition*, *155*, 188-203. doi.org/10.1016/j.cognition.2016.07.003
- Poltrock, S., Chen, H., Kwok, C., Cheung, H., & Nazzi, T. (2018). Adult learning of words in a nonnative language: Consonants, vowels, and tones. *Frontiers in Psychology*, *9*:1211. doi: 10.3389/fpsyg.2018.01211.
- Poltrock, S. & Nazzi, T. (2015). Consonant/vowel asymmetry in early word form recognition. *Journal of Experimental Child Psychology*, *1312*, 135-148. doi.org/10.1016/j.jecp.2014.11.011
- Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, *33*, 3-11. DOI: 10.1037//0012-1649.33.1.3
- Singh, L., Goh, H.H., & Wewalaarachchi, T.D. (2015). Spoken word recognition in early childhood: Comparative effects of vowel, consonant and lexical tone variation. *Cognition*, *142*, 1-11. doi.org/10.1016/j.cognition.2015.05.010
- Stager, C.L. & Werker, J.F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381-382. DOI: 10.1038/41102

- Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76(2), 147-166. DOI: 10.1016/S0010-0277(00)00081-0
- Tardiff, T., & Fletcher, P. (2008). *Chinese Communicative Development Inventories: User's guide and manual (Putonghua and Cantonese versions)*. Beijing, China: Peking University Medical Press. ISBN: 978-7-81116-516-6
- Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, 10(2), 172-175. doi:10.1111/1467-9280.00127
- Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, 17(4), 432-444. doi:10.1111/j.1532-7078.2011.00084.x
- Toro, J.M., Nespors, M., Mehler, J., & Bonatti, L.L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science*, 19, 137-144. DOI: 10.1111/j.1467-9280.2008.02059.x
- van Ommen, S., Boll-Avetisyan, N., Larraza, S., Wellmann, C., Bijeljac-Babic, R., Höhle, B., & Nazzi, T. (2020). Language-specific prosodic acquisition: a comparison of phrase boundary perception by French- and German-learning infants. *Journal of Memory and Language*, 104108. <https://doi.org/10.1016/j.jml.2020.104108>
- van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, 24, 573-583. DOI: 10.3758/BF03201084
- Von Holzen, K. & Nazzi, T. (2020). Emergence of the C-bias during the first year of life: New evidence from own name recognition. *Infancy*. DOI: 10.1111/infa.12331
- Wang, W.S.-Y. (1963). Mandarin phonology. *Project on Linguistic Analysis*, 6, 1-6. ASIN: B00BGLHZ00
- Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63. doi.org/10.1016/S0163-6383(84)80022-3
- Wewalaarachchi, T. & Singh, L. (2020). Vowel, consonant, and tone variation exert asymmetrical effects on spoken word recognition: Evidence from 6-year-old monolingual and bilingual learners of Mandarin. *Journal of Experimental Child Psychology*, 189, 104698. doi.org/10.1016/j.jecp.2019.104698
- Wewalaarachchi, T.D., Wong, L.H., & Singh, L. (2017). Effects of language experience on phonological sensitivities in early childhood: Evidence from bilingual and monolingual toddlers. *Journal of Child Experimental Psychology*, 159, 16-33. doi: 10.1017/S1366728918000597
- Wiener, S. (2020). Second language learners develop non-native lexical processing biases. *Bilingualism: Language and Cognition*, 23(1), 119-130. doi.org/10.1017/S1366728918001165
- Wiener, S. & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and Speech*, 59, 59-82. DOI: 10.1177/0023830915578000

- Wong, P., Cheng, S.T., & Chen, F. (2018). Cantonese tone identification in three temporal cues in quiet, speech-shaped noise and two-talker babble. *Frontiers in Psychology, 9*:1604. doi: 10.3389/fpsyg.2018.01604
- Yeung, H.H., Chen, K.H., & Werker, J.F. (2013). When does native language input reorganize phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language, 68*, 123-139. doi.org/10.1016/j.jml.2012.09.004
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press. doi.org/10.1017/S0952675703004536