Is Categorical Perception for Phonemes Adult-Like by 6 Years of Age? Phoneme Identity and Reaction Time in the Flower Crown Task for Multilingual Children in Singapore

Han Ke, Lei Pan, Beth Ann O’Brien, and Suzy J. Styles

1. Introduction

Singapore has a rich multi-lingual landscape, and inter-generational multilingualism in diverse languages. Singapore’s education system is bilingual, with most classes conducted in English, alongside language classes for other ‘Mother Tongue’ languages. In this context, most Singaporeans are exposed to English and one of Singapore’s official languages (Chinese Mandarin, Malay, Tamil) from early in development, and 90% of young people report being literate in two or more languages (Wu et al., 2020). According to official Census data, 74% of the population are ethnically Chinese, and 35% of households report Mandarin Chinese as the language spoken most at home (Singapore Census Statistics, 2020), making Mandarin the dominant language outside English. For bilingual infants, exposure to two languages from birth – with their dual phonological systems and lexicons – may result in a pattern of language acquisition that is different from their monolingual peers (Hoff et al., 2012), and this may be particularly complex when the phoneme inventories differ between a bilingual’s languages. To date, there is limited evidence about the nature of language development in complex multilingual environments. In order to support educators and parents in Singapore, it is important to understand how auditory perception skills and phoneme categories develop in Singapore’s multilingual context.

In the linguistic literature, we use the concept of phoneme to refer to a category of speech sounds that can be distinguished from others by its function, for example, distinguishing the meanings of words from one another. Sounds can differ from each other in various ways (Ladefoged & Maddieson, 1996), for example, the English words “peach” and “beach” differ only in their initial...
consonants, the voiceless and voiced bilabial stops /p/ and /b/ respectively. In this minimal pair, the phonetic difference in voicing leads to two different words in English, hence /p/ and /b/ are phonemically contrastive. Successfully recognizing, identifying or manipulating phonemes is a fundamental auditory process for language learners. In older children, forming robust acoustic representations of phoneme categories also supports the mapping between letters and sounds, a key element of emerging literacy (Tamási, Wewalaarachchi, and Höhle 2016). In adults, the categorical nature of speech sound perception can be observed in a phoneme identification task (e.g., ‘Is this sound /b/ or /p/?’), when speech sounds on an acoustic continuum show a steep transition in identification (e.g., transition from 100% /b/ identifications to 100% /p/ identifications). This will result in a characteristic s-shaped psychometric function (Lisker and Abramson 1964), in which the ‘threshold’ indicates the category boundary, and the ‘slope’ value indicates the steepness of the s-shape at the threshold. Another hallmark of categorical perception is slower reaction times at the category boundary (Pisoni and Tash 1974).

Phonological perception develops with linguistic experience (Kuhl 2004). In the type of phoneme known as a stop consonant, the vocal tract is closed while the air from the lungs continues to flow, leading to a build-up of pressure at the point of constriction (e.g., the lips for a /b/ sound). When the closure is released, air rushes out suddenly creating a characteristic pressure wave with a sudden onset (high rise time). The voice onset time (VOT) refers to the time that elapses between the release of closure of the vocal tract, and the onset of the voicing for the vowel following (Lasky et al., 1975; Streeter, 1976). As early as 1 month-of-age, infants show discrimination for tokens on a VOT continuum (Eimas, Siqueland, Jusczyk, & Vigorito 1971; Streeter, 1976).

Few studies have investigated how phonemic category perception develops in bilinguals, and the results have been somewhat mixed. Some behavioural studies indicate that bilingual infants’ developing pattern of phoneme perception is similar to monolingual infants. Burns and colleagues (2007) used a visual habituation looking time paradigm to investigate phoneme perception in English-French bilingual infants. They found that at 6-8 months-of-age both bilingual and monolingual infants can discriminate both the native English and native French /ba - pa/ boundaries (at 48ms and 8ms VOT, respectively). By 11-12 months-of-age, English monolingual babies showed alignment to only the English native boundary, while bilingual babies continued to discriminate both French and English native boundaries. Sundara, Polka and Molar (2008)’s work using the same paradigm showed similar findings, 6–8-month-old infants from all groups succeeded in discriminating both French and English /dæ/ syllables from real words, but at 10-12 months-of-age, English monolingual and bilingual infants show equal sensitivity to the English pattern, and English-French bilingual infants continued to discriminate both patterns.

However, other studies suggest bilingual infants have a delay in their phoneme perception development, relative to monolingual infants. One study using the familiarization-preference procedure found that both Spanish-Catalan
bilingual infants and Catalan monolingual infants could discriminate the Catalan /e - ɛ/ contrast in pseudowords 4 months of age, but at 8 months-of-age the groups differ in their sensitivity to this contrast, with Catalan monolingual infants continuing to show the ability and bilingual infants show a U-shaped trajectory – lacking sensitivity at 8 months-of-age, and regaining the ability at around 12 months-of-age (Bosch and Sebastián-Gallés 2003); a follow up study using the same paradigm with /o - u/, /e - u/ contrasts replicated this U-shaped pattern for bilingual Spanish-Catalan infants (Sebastián-Gallés and Bosch 2009).

Two further studies have examined the effect of language dominance on perception of phonetic contrasts in young bilingual. One Event Related Potentials study tested 6- to 9- and 10- to 12-month-old English-Spanish bilingual infants using a double oddball paradigm (Garcia-Sierra et al. 2011). In this study infants were presented with 12ms VOT /ta/ as the standard stimulus, interspersed with 46ms VOT /ta/ deviants (typical for English) and -24ms VOT /da/ deviants (typical for Spanish). Findings suggested that by 10- to 12- months-of-age bilingual infants showed neural discrimination to both native contrasts, and in particular only those infants who had high exposure to one of the languages show an MMN response at this later age. Furthermore, a behavioural double oddball study also reported an effect of language dominance in 8- to 9-, 11- to 12- and 14- to 15-month-old Dutch monolingual and bilingual infants (Liu & Kager, 2015). In this study, monolingual Dutch infants and bilingual Dutch infants with Dutch as their dominant language showed the same pattern of discrimination. They both showed consistent sensitivity to the native long-lead /b - p/ contrast but not the non-native /p - pʰ/ contrast across the testing age. By contrast, the bilingual Dutch infants with a non-Dutch dominant language showed the opposite pattern, but robust discrimination of their native /p - pʰ/ contrast was only found at a later stage (11-12 months-of-age). These findings showed that the stabilization of contrast discrimination started late for bilingual infants, and that language exposure and language dominance strongly influence bilingual infant’s VOT perceptual patterns.

In more recent investigations, evidence from a study of Korean adoptees’ native language retention (Choi et al., 2017) has suggested that early exposure to a language influences phonological learning in adulthood, even if the early exposure happened before infants consolidate their perceptual phoneme categories. In a similar vein, Hokkien-reared adults in Singapore who have forgotten their caregiving language are significantly faster at learning Hokkien tones compared to English-reared adults with no experience of Hokkien tones (Singh & Seet, 2019). These findings indicate that phonological knowledge is processed even before the 6- to 12-month-old phoneme tuning window, and that language input in early childhood yields an impact on later phonological perception.

To date only Liu and Kager’s study included bilingual infants with a Chinese language background, in which they found Dutch-Chinese bilingual infants showed a delay in discriminating phonetic contrasts in their dominant language. They suggested that this occurred because bilingual infants’ complex environment
brought different input frequencies and distributional properties of the languages. However, a single study with a small number of participants in this group gives limited support for this interpretation. Furthermore, since bilingualism is the majority pattern of language exposure in Singapore, these findings may not be directly relevant to the children’s language development in the local context.

Few studies have investigated how categorical perception of phonemes progresses in bilinguals after the toddler years. For children approaching school-age, the steepness/shallowness of the slope at the threshold has previously been linked to reading difficulties (Maassen et al. 2001; Chiappe, Chiappe, and Siegel 2001; Chiappe, Chiappe, and Gottardo 2004; Boets et al. 2011; Messaoud-Galus, Hazan, and Rosen 2011). However, apart from one recent study showing gradual sharpening of category boundaries continues into adolescence (McMurray et al. 2018), limited literature has addressed the development of phoneme perception skills across childhood, nor the development of these skills in bilingual children. Hence in this study, we focused on how phoneme perception differs for bilingual children and adults in the Singaporean bilingual population, in terms of threshold, slope and reaction time, in a phoneme identification task.

One additional feature for consideration is the unique distributional properties of Singapore’s languages, arising from the long history of language contact in the region. Acoustic measurements of speech have shown that Singapore English (Huang, 2003) and Singapore Mandarin (Ng, 2005) have quite different VOTs for their stop consonants, with typical voiced stops /b, d, g/ and voiceless stops /p, t, k/ occurring 9ms and 90ms for Singapore Mandarin, and 4ms and 25ms for Singapore English (with notable pre-voicing in certain contexts). Other common languages in Singapore have different acoustic patterns, such as Singapore Malay: 11ms VOT for /b, d, g/ and 16ms VOT for /p, t, k/; and Singapore Hokkien: 8ms VOT for /b, d, g/ and 77ms for /p, t, k/. These are all somewhat different from the VOTs of British English (21ms and 56ms). With such a variety of VOT exposure in the local language landscape, it is necessary to design acoustic paradigms specifically for Singaporean children to reveal their language development pattern.

In the current study, we aimed to investigate the development of children’s phonetic categories in the early school years using a phonological perception task designed specifically for the Singapore language environment. The phonetic category perception task is based on Lisker and Abramson’s method of phoneme identification (Lisker & Abramson, 1970; Abramson & Lisker, 1970; 1973). In this preliminary report, we present data from bilingual Chinese/English speaking adults and Chinese/English speaking children. The data presented here are related to two pre-registered analysis plans, currently in progress (Adults: Pan, Ke, & Styles, 2020; Children: O’Brien et al., 2020). Detailed comparisons will be conducted following the preregistered stop dates for each study.
2. Methods
2.1. Participants

We recruited 69 Singaporean adults in a local university, and 127 children in their second year of kindergarten at preschools and kindergartens in Singapore that have a sizable proportion of students enrolled in Chinese mother tongue language classes and thus have bilingual English and Chinese backgrounds. Schools helped distribute recruitment flyers and consent forms which parents could opt into by signing in advance of a test conducted at the School, or in a developmental testing centre. Adult participants were recruited from the University population, and gave consent using a consent form. Children and adults participated in the same assessments for the same duration.

3 adults and 12 children were excluded from final analysis due to wrong language background, or failing to complete the task correctly. The final sample included in this preliminary analysis is 66 adults ($F = 51$; Age: $M = 21.1$, $SD = 1.9$) and 115 children ($F = 54$; Age: $M = 6.3$ years, $SD = 0.3$ years).

2.2. Paradigm

Participants were asked to complete the Flower Crown task. This task asked participants to make a simple decision about which of two pictures matches a spoken word, in the guise of a game where participants help a little monkey to collect flowers for his crown. Participants performed the decision several times for speech sounds along a VOT continuum. We constructed a series of audio stimuli ranging across a Voice onset time (VOT) continuum ranging from very pre-voiced (-60ms) to very aspirated (+90ms), in 10 milliseconds steps from /beach/ to /peach/. The audio segments were created by recording a speaker of the local variety of Singapore English articulating two speech tokens (‘beach’ and ‘peach’). The VOT spectrum was created by splicing the audio in steps of 10ms from minus 60ms (pre-voicing) to plus 90ms (aspiration), resulting in 16 tokens.

To evaluate the category structure for each individual, during the Flower Crown task, two images were presented on either side of the screen. Participants heard an audio file drawn randomly from the VOT continuum. Participants heard one of the tokens of speech and were asked to select the matching picture, as shown in Figure 1. The presentation software was OpenSesame (python 2.7 version, Ke, Pan & Styles, 2020). Over the 10 blocks of the trial, participants made 10 decisions for each step on the VOT continuum, allowing a psychometric function to be fitted for each individual. The psychometric function allows the measure of the threshold and slope to be derived for each individual. The task also recorded the reaction time of each decision (see Variables section for details of measurements). Curve fitting was conducted on proportion of /p/ responses using the quickpsy package in R (Linares & López-Moliner, 2016; R Core Team, 2013), and rescaled to percentage for graphing. The code for running the open-source procedure can be found in the OSF repository for this project (Ke, Pan & Styles, 2020).
2.3. Variables

To conduct the analysis, the following variables were recorded and/or computed.

2.3.1. Threshold

At each VOT on the continuum, the percentage of /p/ decisions were computed, and an individual identification curve across the continuum was plotted. For each individual, the binary decisions of /b/ and /p/ at different steps on the continuum were fitted to a logistic function, as a way of ‘smoothing’ responses that may contain noise/errors, following the statistical procedure of Boets et al., (2011).

The threshold is the point on the fitted psychometric function at which an individual switches from making majority ‘beach’ decisions to making majority ‘peach’ decisions (i.e., the VOT at 50% /b/ 50% /p/ decisions). The threshold appeared at different places on the VOT continuum depending on persistent individual differences, or language exposure (i.e., what proportion of a child’s input is English versus Mandarin).

2.3.2. Slope

The slope value for each individual was derived from the logistic regression of their responses across the spectrum (rescaled for graphing).

2.3.3. Prototypicality

We also computed a measure of Prototypicality at each VOT as the absolute value of the percentage of choices for a particular phoneme, centered on a zero axis. This resulted in a Prototypicality of 1 when a participant showed 100% of responses for the same phoneme (i.e., all /b/ choices OR all /p/ choices), and a
Prototypicality of 0 at the VOT boundary for an individual (i.e., 50% /b/ choices, 50% /p/ choices).

### 2.3.4. Reaction time (RT)

RTs were recorded in each trial from the beginning of the audio stimulus presentation until the participant pressed the response button. See the analysis section for the details of RT handling.

### 2.4. Analysis

We predicted that bilingual adults and children in Singapore would show a range of responses to the Flower Crown Task, both in the location of their category boundaries, and the in steepness of the slope between categories; and that adults’ reaction times (RTs) would differ across the VOT spectrum, in particular, the RTs would be slower when the VOT was closer to their crossover point, due to the prototype effect (Pisoni & Tash, 1974; Massaro, 1989). That is to say, participants would show the slowest reaction times for items near their individual category boundary, and fastest for items that are more prototypical. However, the children’s RTs might not show this pattern yet. To test the hypothesis, for each age group we evaluated the influence of Prototypicality at each VOT step as a fixed effect on the average of the transformed RT at each time step, using a generalized linear mixed model with participant as a Random effect.

Due to the nature of RTs, we transformed the reaction time values before analysis. As RT distributions were rarely normal (i.e. non-Gaussian), we checked for normality by plotting histograms of the RTs. If the histogram was positively skewed with a long tail, the data was then handled by transformation to approximate normality. First, RTs were transformed using both log and inverse function, after visual inspection, the transformation with a histogram that best approximates normality for the majority of the dataset was selected. The histogram was checked visually for extreme outliers (e.g., those separated from the majority by a large gap). Extreme outliers were trimmed. The data was determined to be sufficiently close to normal using the reported methods above, the parametric tests described above were then used for analysis. Analysis software was R (version 4.0.3, R Core Team, 2013).

### 3. Results

#### 3.1. Threshold and slope

Figure 2 shows the fitted curves for each child in the study, and each adult in the comparison group. We observed children’s VOT thresholds were slightly later than adults (Children: $M = 34$ms; Range = -51ms – 174ms; Adults: $M = 26$ms, Range = 12ms – 46ms). We also found that children’s VOT perception curves had a substantially shallower slope than those of adult bilinguals from the same speech
community, with almost no overlap between groups (Children: $M = 5.21$, Range $= -0.55 - 21.29$; Adult: $M = 69.72$, Range $= 4.35 - 865.60$).

Figure 2. Fitted identity curves at each VOT step for children (left) and adults (right), with the median participant in each group highlighted.

Figure 3 shows individual threshold and slope for the two groups. Note that a large slope value indicates a steep slope, as can be seen in Figure 2.

Figure 3. Marginal distribution of slope values and threshold of children (left) and adults (right)
3.2. Reaction times

For each age group, we conducted general linear mixed model with the log transformed RTs as dependent variable, and the prototypicality of percentage of /p/ as predictor, using binomial method. Results revealed that adults’ reaction time showed a typical pattern of slower RTs near the crossover (tokens with low prototypicality for the individual) ($X^2 (1) = 212.92s$, $p < .001$). The mean RTs at the prototypicality of 1 were 546ms ($SD = 128ms$), while the mean RTs at the prototypicality of 0 was 661ms ($SD = 274ms$). However, children’s reaction time does not change significantly with the VOT spectrum, $X^2(1) = 0.25$, $p = .62$. The mean RTs at the prototypicality of 1 were 1054ms ($SD = 537ms$), while the mean RTs at the prototypicality of 0 was 994ms ($SD = 721ms$). See Figure 4 and 5 for the individual RTs spread at each VOT step, and at each prototypicality point. See Figure 4 and 5 for the individual RTs spread at each VOT step, and at each prototypicality point.

Figure 4. Violin plot of the mean of the RTs per individual at each VOT step for children (left) and adults (right), with the median value of the group highlighted

Figure 5. Jittered scatter plots of reaction time and prototypicality for children (left) and adults (right)
4. Discussion

In this study, we investigated phonological perception in Singaporean children and adults from the same Chinese-English bilingual language background. Results revealed that at 6 years of age bilingual children showed a characteristic S-curve indicating categorical perception of English phoneme /b/ and /p/ in a child-friendly picture identification game. Compared to the adults from the same community, the spread of the responses was quite large, and the average slope value was shallower, indicating more variability between individuals, and on average, less clear boundaries between categories.

Overall, children showed slower reaction times compared to the adult comparison group. Only adults showed a significant slow-down at their individual category boundary – a hallmark of strong categorical perception. Preliminary analyses reported elsewhere suggested that threshold and slope were not influenced by the precise balance of the language mix in children of this age (poster: Ke et al., 2020). In general children and adults showed similar thresholds for the perceptual boundary between /b/ and /p/ in English. Children who deviated from this pattern also exhibited extreme slope values, suggesting atypical categorical perception for this contrast. This suggests a combination of slope value and threshold maybe useful for identifying children at risk of developing language delay/disorders.

In summary, by the age of 6 years, Singaporean English/Chinese bilingual children’s perception of English phonemes /b/ and /p/ shows shallower categorical perception at the same acoustic boundary as adults, with less entrenched advantage for prototypes over boundary exemplars. Phoneme development appears to be not yet fully mature at this age.

References


Ke, Han, Pan, Lei, and Styles, Suzy J. 2020. Flower Crown Task: An open access phoneme identification task. https://doi.org/10.17605/OSF.IO/F3B4C


