

Summary

Perception of Vowels in Sequential Bilinguals

Doğu Erdener

Middle East Technical University
Northern Cyprus Campus

Denis Burnham

Western Sydney University

It has been known since the early 1970s that within the first year of life infants can discriminate, most, if not all, speech contrasts across the world's languages in a broad, language-general, phonetically/acoustically-based fashion (e.g., Best, McRoberts, & Sithole, 1988; Burnham, 1986). However, this early phonetically-based ability is reorganized during the first year of life such that a native language (L1) bias occurs first for vowels sometime between 4 and 6 months of age (Kuhl, Williams, Lacerda, & Stevens, 1992) followed by the reorganization of consonants (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984) and lexical tones (Mattock & Burnham, 2006) between 7 and 11 months. This reorganization process appears to be a result of specific linguistic experience with L1 - a transition from language-free, phonetically based processing to a language-specific, language-specific one. An extensively investigated example of this reorganization is the perception of /r/ vs. /l/ contrast by native Japanese speakers. While it was found that adult Japanese speakers have difficulty in discriminating this contrast (Goto, 1971; Hardison, 2003; Logan, Lively, & Pisoni, 1991; Sheldon & Strange, 1982), their infant counterparts can do this with ease (Kuhl, 1997).

Interestingly, some nonnative speech contrasts can be discriminated by adults on the basis of their acoustic rather than phonological properties such as Zulu clicks (Best et al., 1988). Best (1995), in her Perceptual Assimilation Model (PAM), suggests that the way one can discriminate nonnative speech contrast is a function of the degree to which that contrast can be assimilated into a native phonemic category, which, apparently, occurs via three ways: (a) a nonnative speech sound can be assimilated into a native phonemic category; (b) a nonnative phoneme may be categorized as a nonnative speech sound; (c) a nonnative speech sound may be identified as a non-speech sound (e.g., Zulu clicks, see Best et al., 1988). For instance, according to this model, Japanese

adults' inability to discriminate /r/ vs. /l/ contrast can be explained by the fact that these two phonemes are mapped onto the single flap /r/ in the Japanese phonemic repertoire leading to the assimilation of these two phonemes into the Japanese-native phonemic category of /r/.

Research shows that learning a second language (L2) affects the way we perceive speech; however, whether speaking more than one language improves speech perception is a matter of debate. In his Second Language Model (SLM), Flege suggests that factors such as the relative use of L1 and L2 (Flege, 1995; Flege, Frieda, & Nozawa, 1997; Flege, Munro, & MacKay, 1995; Guion, Flege, & Loftin, 2000), age of acquisition (Flege et al., 1995), and training (Flege, Takagi, & Mann, 1995) affect the way L2 speech production/accent as perceived by native speakers of that L1.

Studies that pertain to the impact of speaking more than one language on nonnative speech perception yield mixed results at first sight. While Werker (1986) found no difference amongst three groups of monolinguals, bilinguals and trilinguals on the discrimination of Hindi retroflex/dental and Nthlakampx glottalized velar/glottalized uvular phonetic contrasts, Greek / English bilingual adults were found to be marginally better than their English-speaking monolingual counterparts when tested on their perception and production of consonant-vowel (CV) Thai speech stimuli (Beach, Burnham, & Kitamura, 2001).

Bilingualism has been investigated in a plethora of aspects; however, relatively the least attention has probably been paid to the phonological side. This study aims to investigate the perception of nonnative vowels by early and late Turkish / English bilingual speakers. Vowels are particularly important as they have unique characteristics linguistically and developmentally. Developmentally, similar to consonants and lexical tones, the perception of vowels shows a language-general to language-specific developmental pattern, yet this occurs earlier for vowels (Kuhl et al., 1992; Polka & Werker, 1994).

In this study, we decided to recruit two bilingual groups on the basis of whether they were born in Australia and learnt English since birth and prior to school age *or* whether they arrived in Australia after the school age and learnt English after their Turkish was well established. These definitions *also* factored in whether the L2 was acquired before or after reading acquisition in L1 (and L2). It was hypothesized that if early bilinguals have an advantage as they presumably formed their mental representations of phoneme categories in English and Turkish before reading acquisition, then they should be able to discriminate target stimuli better than their late bilingual counterparts. In addition, as early bilinguals had pre-reading exposure to two phonological systems (Turkish and English) and if this provides them with a phonological space wherein the phonemic categories belonging to the two language are clearly defined, then they should also perform better than their monolingual counterparts who were exposed only to a single phoneme system (English).

Method

There were three speaker groups: early bilinguals, late bilinguals, and monolinguals. A total of 42 (19 males and 23 females) participants were recruited for the study, and each speaker group contained 14 participants. The mean age of the participants was 28 with a range of 38 years. The participants were recruited via word of mouth at the Universities of New South Wales and Sydney, as well as public radio announcements. The criteria set for the speaker groups were as follows: the *early bilinguals* were required to have been born in and lived continuously in Australia, or migrated to Australia before the age of four, (prior to the onset of schooling), be able to speak and understand English and Turkish before they started school, and have had no prior experience with a tonal language (e.g., Mandarin, Cantonese, Thai).

For the *late bilinguals*, the criteria set by Tyler (2001) were adopted. The late bilingual participants were required to have started learning English after the age of 10, immigrated to Australia after the age of 15, been living in Australia for five years or more and had no experience with a tonal language.

Materials and Stimuli

Using five vowels, four vowel contrasts were created: /ɔ/, /u/, /ə/, /u/, and /o/. Participants were familiarized with the AX discrimination task with two “same” and two “different” trials consisting of the words “rag” and “rug”. The stimuli were recorded with the respective native speakers of all three stimulus languages – three female speakers and edited using the CSL software (see

<http://fastlinabinc.com/CSL> for an updated version). The vowel components were isolated via a home brand software named CSL2BIN developed at the University of New South Wales (UNSW) in Sydney, Australia and rendered ready for presentation via another software MAKEDIS. The average duration of each vowel stimulus was set at 60-70 milliseconds. Responses were collected with a USB response box attached to the laptop computer. The stimuli were presented via a professional headset (OPTIMUS NOVA44).

Procedure

Testing sessions were conducted individually in a quiet testing room at UNSW’s psychology department. Some participants were tested in their homes due to their inability to travel for lack of time. It was ensured that the rooms used testing did not have a noise level of over 45 dB. The experiment consisted of two sections. The first one was the familiarization task as depicted above. The second phase, namely the experimental stage, was made up of 72 items with two inter-stimulus interval (ISI) versions: 500-msec and 1500-msec. Each ISI level was presented as a separate block and repeated twice with block orders counterbalanced and stimuli randomly presented within the each block. The 500-msec stimuli aimed at measuring the discrimination ability at the phonetic level as this period of time does not allow for language-specific, phonological processing, while the 1500-msec ISI stimuli were designed for phonologically-based discrimination.

The dependent variable was a discrimination index (DI) score which was obtained via dividing the difference between correct responses to the “different” items and incorrect responses to the “same” items by the total number of “different” items. The advantage of this formula was that it factors in the random “different” responses to the “same” trials. DI scores ranged from -1 to +1, reflecting poor and superior performances, respectively. A “zero” performance revealed a chance-level responding.

Results

Given the small sizes of the three participant groups, ($N = 42$, each $n = 14$), Kruskal-Wallis H test was used as a parametric measure in lieu of analysis of variance. None of the 500-msec ISI analyses reached significance except the one for Thai-only contrast performance, $\chi^2(2, N = 42) = 6.365, p = .041$. Three further Mann-Whitney U tests were conducted as post-hoc analyses comparing three groups over their performances on Thai-only stimuli. The only significant difference was between early and late bilinguals, $z(N = 28) = -2.456, p = .01, r^2 = .01$

and the other two between-group comparisons (late bilinguals vs. monolinguals and early bilinguals vs. monolinguals) reached no significance.

While none of the Kruskal-Wallis H tests for the 1500-msec stimuli revealed any significance, akin to the 500-msec ISI results, Thai-only contrasts were very close to the significance, $z(2, N = 42) = 6.365, p = .056$. Given the small sample size, we took the liberty of running three Mann-Whitney U tests over the three groups' scores on Thai-only contrasts in lieu of post-hoc analyses. The results were identical to those found for 500-msec ISI condition such that the only significant difference was the one between early and late bilingual groups, $z(N = 28) = -2.200, p = .03, r^2 = .03$.

Discussion

This study predicted that early exposure to larger numbers of vowel phoneme categories would lead to better perception of these elements in bilinguals, particularly those with earlier exposure (i.e., early bilinguals). The general prediction was such that early bilinguals would have surpassed the other two groups in most, if not all, contrasts. However, this was the case only in the Thai-only contrasts which were non-native to all participant groups such that in both phonetic (500-msec ISI condition) and, albeit marginally ($p = .056$), phonological conditions, early bilinguals discriminated Thai-only contrasts superiorly. Although at non-significant levels (and considering the small group sizes), what was surprising was the finding that monolinguals performed better than their late bilingual counterparts in all stimulus conditions and in both ISI conditions. Furthermore, early bilinguals performed better than their late counterparts but not better than monolinguals over these Thai-only stimuli. These results pinpoint two paths. First, the acquisition of L2 occurring prior to literacy skills somehow paves the way for the formation of phonemic categories that have higher sensitivity (Burnham, 2003; Horlyck et al., 2011). Whether or not early bilinguals end up with greater number of phonemic categories than late bilinguals or monolinguals cannot be determined here. However, what is interesting here is that early bilinguals seem to

be more sensitive to vowel contrasts that are not present in either of their languages. The fact that this occurs in early bilinguals who acquired their respective L1 and L2 prior to reading (and at both phonetic and phonological levels) lends support to the following notion: pre-reading acquisition of L2 allows for the formation of some sort of sensitivity most likely stemming from greater number of vowel categories. Besides, Burnham's (2003) finding which suggests that speech perception is further altered or reorganized as a result of exposure to literacy skills also lends partial support to the differences observed between early bilinguals and the other two groups given that early bilinguals' L2 acquisition was before reading acquisition.

In conclusion, we have two paths of explanation for results obtained here. First, earlier exposure does not solely operate at language-specific, phonological level, but at phonetic level, encompassing two levels of speech processing. Secondly, it was demonstrated here that previously unexposed vowel stimuli (Thai) are perceived at more shallow structures such as phonetic level. Whilst developmental studies show that phonetic processes are followed by phonological processes, this research, at least on the basis of non-native vowel contrasts, showed that these two processes work in a parallel and integrated fashion (see the results from 500-msec vs 1500-msec stimuli).

This is one of the hypotheses that can be advanced as to the roles of phonetic and phonological processes that may be in operation during L2 acquisition onset at different phases of life. Further to this, there is a growing body of research which now clearly suggests that speech perception is not solely an auditory phenomenon but an auditory-visual one in the sense that we use orofacial (mouth and lip) movements to decipher incoming speech signal. Research on both auditory and visual aspects of speech and their integration in the context of L2 acquisition, is in fact, a growing body of research as well yet there is some paucity of research, particularly in L2 context involving non-English languages. Adoption of such a perspective of research and openness in the practice to utilize the findings of this research will pave the way for a more in-depth understanding of L2 acquisition (Erden, 2016).

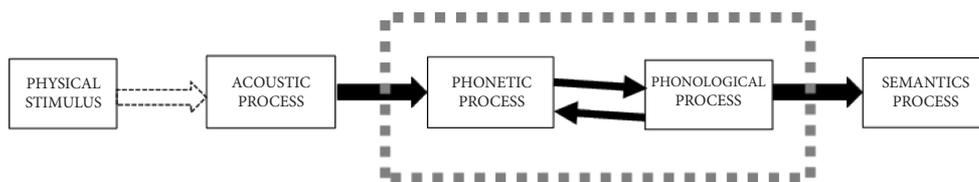


Figure E1. Phonetic and Phonological Processes in Speech Perception