Learning challenging L2 sounds via computer training:

High-variability perceptual training for children and adults

Kathy Kar-man Shum[1*]

Terry Kit-fong Au[1]

Laura F. Romo[2]

Sun-Ah Jun[3]

[1] Department of Psychology, The University of Hong Kong, Hong Kong, China

[2] The Gevirtz Graduate School of Education, University of California Santa Barbara, Santa Barbara, United States of America

[3] Department of Linguistics, University of California Los Angeles, Los Angeles, United States of America

* Corresponding author

E-mail: kkmshum@hku.hk

Learning Challenging L2 Sounds via Computer Training: High-Variability Perceptual Training for Children and Adults

Abstract

Do learners of a second language (L2) need frequent contact with native speakers of that language in order to master its phonology? What if they hear audio recordings of native speakers and receive immediate corrective feedback about their perception? We used a randomized controlled experiment with 135 Chinese speakers (with English as their L2) to examine whether a high-variability perceptual training (HVPT) paradigm might enhance the perception of challenging contrasts between English voiced and voiceless stop consonants. Learners in all the age groups tested—middle childhood, early adolescence, and young adulthood—showed enhanced perception of English stop consonants after 20 five-minute training sessions conducted across 4 to 6 weeks, based on audio-recorded input coupled with corrective feedback. The training benefits were maintained at the one-month follow-up. Our results suggest that HVPT using audio-recordings of native speakers can be an affordable and useful language enrichment to supplement live interaction with native speakers, for L2 learners of a wide age range.

Keywords: high-variability perceptual training; second language; phonology; input; consonants

Learning Challenging L2 Sounds via Computer Training: High-Variability Perceptual Training for Children and Adults

Introduction

With globalization, learning a second language (L2) has become commonplace. Finding effective ways to support L2 learning has become an educational priority more than ever. Linguistic input from native speakers is one key, which seems especially crucial for enhancing the phonology and morphosyntax among L2 learners, more so than for some other aspects of language such as vocabulary and basic word order (Curtiss, 2014; Johnson & Newport, 1989). It would sound wise to offer L2 learners linguistic input from native speakers to help them master their L2 phonology (Flege, Yeni-Komshian, & Liu, 1999; Oyama, 1976). Unfortunately, when resources are limited, and when the L2 is not the societal language (e.g., children learning English in much of Africa, Asia, and Latin America), live interaction with native L2 speakers is often scarcely available. Audio recordings are commonly used as an affordable substitute. Can this type of training actually help L2 learners acquire better L2 phonology?

There is some promising evidence that merely hearing audio recordings of songs and stories in an L2 can improve young children's L2 accents (Au, 2013; Au, Chan, Cheng, Siegel, & Tso, 2015). Intensive high-variability perceptual training (HVPT) with corrective feedback has also been shown to help adult learners learn the contrasts between basic speech sounds in L2. For example, Japanese adults improved in both their perception and production of the English /r/ and /l/ contrast, and maintained the improvement in a 3-month follow-up (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002). High-variability perceptual

training was also shown to improve the perception of English vowels among Chinese Mandarin speakers (Thomson, 2012), as well as vowel perception and production among native French speakers learning English as L2 (Iverson, Pinet, & Evans, 2012).

To date, relatively few studies have examined the effects of HVPT among children (Giannakopoulou, Brown, Clayards, & Wonnacott, 2017; Giannakopoulou, Uther, & Ylinen, 2013; Shinohara, 2014; Shinohara & Iverson, 2013). HVPT has been shown to help Japanese children and adolescents improve their English /r/-/l/ identification on the trained word-initial position, although not in untrained positions such as medial and consonant clusters (Shinohara, 2014; Shinohara & Iverson, 2013). Giannakopoulou et al. (2013, 2017) also found improvement in the perception of the English /ɪː/-/ɪ/ (tense-lax) vowel distinction in 7- to 8-year-old Greek L2 speakers after HVPT, and training benefits were generalized to untrained words. These studies provided evidence to support the feasibility and effectiveness of HVPT in enhancing children's phonetic learning.

It would certainly be of great interest to researchers and practitioners alike to compare the effects of HVPT on children in contrast to adults. On the one hand, adults may benefit more from intensive perceptual training due to their better attention control and better allocation of cognitive resources (Antoniou & Wong, 2015). It is noteworthy that stimulus variability in the training paradigm exposes the learner to a variety of exemplars of the target contrast, and is thought to lead to more robust categorical learning and better generalization to novel stimuli (Lively, Logan, & Pisoni, 1993). The inclusion of talker variability in HVPT may enhance learning outcomes via more effortful processing on the learners' part during the training, and hence results in more superior long-term retention of phonetic information (Barcroft & Sommers, 2005). The additional processing costs may, however, hinder the learning of low-

aptitude individuals who have fewer cognitive resources to handle the cognitive load (Antoniou & Wong, 2015). This could well be the case among children who may have fewer available cognitive resources or are poorer at keeping their attention on training stimuli.

Nonetheless from another perspective, children might outperform adults in terms of L2 training benefits, possibly due to enhanced brain plasticity (Kuhl, 2004). Much research has reported declines in language learning capacities with age for L2 as well as L1 (Johnson & Newport, 1989; Kuhl, 2004; Newport, 1990), purportedly related to maturing non-linguistic cognitive abilities (Newport, 1990), and increasing neural commitment to structures necessary for L1 development (Kuhl, 2004). The sensitive period view suggests that language acquisition is best begun early in childhood (Curtiss, 2014; Johnson & Newport, 1989), especially for phonology (Flege et al., 1999; Oyama, 1976).

Hence, given the two sides of arguments regarding the speech training effects for different age groups, it is unclear then how robust the benefits of HVPT for younger learners will turn out to be as compared to adults. Thus far, the handful of HVPT studies comparing children and adults have produced mixed results (Giannakopoulou et al., 2013, 2017; Shinohara, 2014). Comparing Greek 7- to 8-year-olds with adults (aged 20 to 30 years), Giannakopoulou et al. (2013) observed more pronounced improvement in English vowel discrimination in children than adults, but this was not replicated using a similar training paradigm (Giannakopoulou et al., 2017). Shinohara (2014) found that Japanese adolescents (aged 15 to 18 years) and older children (aged 8 to 12 years) improved more than either younger children (aged 6 to 8 years) or adults. Although the brain plasticity account might explain the training advantage seen in the older children and adolescents over the adults (Shinohara, 2014), it failed to elucidate the lesser learning in the younger children compared with older age groups. Other factors, such as the

length of training, the salience of the target contrasts to L2 learners, as well as the maturity level of children's phonemic awareness and cognitive abilities including selective attention have been suggested to influence the effects of HVPT on children versus adults (Giannakopoulou et al., 2017; Shinohara, 2014).

We are mindful that traditional HVPT methods are very repetitive and likely boring to the learners—especially children who typically have less discipline and attention control than adults. Some HVPT studies involving child participants have reported adding cartoon animations in the computer program to motivate the children in particular. For instance, Giannakopoulou et al. (2013) used happy and sad animations to provide feedback during training on correct and incorrect responses respectively; Shinohara (2014) used two cartoon characters each corresponding to either /r/ or /l/ words to help the participants to remember the two phonemes better. To motivate the younger participants in our study, we incorporated game-like features to the HVPT in hopes of making the training more engaging, and examined if children as well as adults can benefit from HVPT robustly.

Chinese learning L2 phonology: English stop consonants

We focused on Hong Kong Chinese learning English stop consonants because stop consonants are ubiquitous in English and they are very challenging for native Chinese speakers to master (this is further discussed later). In English, there is systematic contrast in stop consonants between voiced stops /b, d, g/ and voiceless stops /p, t, k/. Here we focused on American English mainly because it has become increasingly popular among L2 learners in China and across the globe, perhaps overtaking British English (Conrad & Rubal-Lopez, 2011), and secondly because our bi-national research team was based in California and Hong Kong.

To begin with, the stop-consonant phonology in English is complicated. Voiced stops /b, d, g/ and voiceless stops /p, t, k/ have different allophones depending on their location in a word. There are voiceless aspirated stops [p^h, t^h, k^h], voiceless unaspirated stops [p, t, k], and voiced stops [b, d, g]. For the alveolar stops /t, d/ in American English, there is also an additional allophone, a flap [ɾ].

In word-initial position, English voiceless stops are aspirated [p^h, t^h, k^h] (e.g., ***p****in,* ***t****ip,* ***c****oat*) with a long-lag voice onset time (VOT; the interval between stop release and the onset of vocal cord vibration), while word-initial voiced stops typically have a short-lag VOT (Lieberman & Blumstein, 1988), especially when the word is utterance initial. Hence, word-initial voiced stops are often realized as what would be perceived as voiceless unaspirated [p, t, k] in other languages (e.g., ***b****in,* ***d****ip,* ***g****oat*). Therefore, the contrast is largely cued by the presence or absence of aspiration (the amount of air released after the stop closure until the onset of the following vowel).

In word-final position where the stop is a coda consonant, the contrast for the two sets of stop consonants (i.e., /p, t, k/ versus /b, d, g/) is no longer cued primarily by aspiration but by the voicing during the stop closure (e.g., presence of vocal cord vibration during the closure), along with the length of the vowel before the stop (Kluender, Diehl, & Wright, 1988). Typically, vowels are longer before voiced stops than before voiceless stops (Chen, 1970; Raphael, 1972; Walsh & Parker, 1981). Voiceless stops in word-final position can be released (aspirated) or unreleased (unaspirated) when pronounced, and thus aspiration contrast cannot help to distinguish voiced versus voiceless coda stops when they are unreleased. Hence, voicing and vowel duration—instead of aspiration—serve as primary cues for the stop consonants in word-final position.

In word-medial intervocalic position where the stop is ambisyllabic (e.g., *co**pp**er, ni**bb**le*; Kahn, 2015), the voicing contrast is cued by the voicing during the stop closure as in word-final position, but not much by the length of the preceding vowel (Lisker & Abramson, 1964). Similar voicing contrasts are also found in word-final stops followed by an unstressed vowel (i.e., *ca**p** it, mo**b** it*; Flege, Munro, & MacKay, 1996). Finally, when a voiceless stop occurs after a tautosyllabic /s/ (e.g., *s**p**in, s**t**op, s**k**ate*), the voiceless stops are unaspirated. That is, voiceless stops are realized as voiceless unaspirated stops [p, t, k] when the stop is in word medial or final position or when it follows a tautosyllabic /s/, while voiced stops are also realized as voiceless unaspirated stop when it is in word-initial position.

Chinese speakers learning English-as-L2 have to deal with L1-related potential interference (Flege & Wang, 1989). There are two sets of stop consonants in Chinese (the unaspirated /p, t, k/ and the aspirated /p^h, t^h, k^h/) that are both voiceless and only differ by aspiration. There are no voiced stops in Mandarin, Cantonese, or indeed virtually all Chinese dialects; all stops in Chinese—whether unaspirated or aspirated—are realized as voiceless in all positions in a word. As noted earlier, word-initial stops in English are largely contrasted by aspiration rather than voicing. The aspiration contrast should be easy to discriminate for Chinese speakers learning English-as-L2, as this is a familiar contrast commonly encountered in their L1 Chinese. In fact, Deterding and Nolan (2007) showed that the average duration of aspiration for word-initial aspirated and unaspirated stops was not significantly different between Chinese and English speakers. Given that Chinese speakers should be good at distinguishing English voiced and voiceless stops in the word-initial position as the contrast is primarily in aspiration rather than voicing, word-initial stops were excluded from the HVPT in the current study.

On the other hand, Chinese speakers who have not received regular input from native English speakers generally do not master the contrast between the English /b, d, g/ and /p, t, k/ in positions other than word-initial, even after more than a decade of English-as-L2 education. For example, Au et al. (2017) showed that only 1 out of 90 Chinese students at a university with the highest admission standard for English abilities in Hong Kong were able to pronounce the voiced stop /b/ correctly in the word "*cub*" for all three times in a scripted conversation. Moreover, among the 270 speech tokens, 88% were judged by one or more of three native English speakers to be mispronunciations—mostly as "cup" (i.e., without lengthening the vowel and/or voicing the /b/; Au et al., 2017). Indeed, it has been shown that all voiced stops in English are generally produced by Chinese speakers as voiceless unaspirated stops in their L2 English, and the vowel duration before a stop is either not manipulated or only minimally (Flege, 1988; Flege, Munro, & Skelton, 1992). Therefore, voicing contrast in stops should be hard to acquire by Chinese speakers but aspiration contrast should be easy. In other words, Chinese learners of L2 English should be better at producing and perceiving word-initial stops in English than word-medial or final stops in English.

While English-as-L2 begins for almost all children attending kindergarten in Hong Kong and continues throughout elementary and high school, most Hong Kong Chinese children—except those at international schools—do not get regular input from native English speakers because of educational resource constraints. These English L2 learners, therefore, offer a valuable window onto how well learners of different ages can acquire the basic but challenging contrast between /p, t, k/ and /b, d, g/ using audio-recorded input from native speakers. Using a randomized controlled experimental design, we tested these hypotheses:

H1: HVPT with immediate corrective feedback improves Chinese speakers' perception of stop consonants in L2 English.

H2: HVPT benefits both adults and children.

H3: Training benefits of HVPT will maintain (e.g., lasting for at least one month after training).

H4: HVPT improves perception of English stop consonants in the word-medial and word-final positions, but not in the word-initial position for which Chinese speakers should be good at prior to the training.

Method

In a randomized controlled experiment, we evaluated how well Chinese speakers—children, early adolescents, and young adults—acquired the challenging contrast between /p, t, k/ and /b, d, g/ in English when they heard identical audio-recorded input from four native English speakers and received immediate corrective feedback on their word perception. The participants in each age group were randomly assigned to either perceptual training focusing on the English stop consonants /p, t, k, b, d, g/ or the control group. Both groups took a baseline pretest (Time 1) and a posttest 4 to 6 weeks later (Time 2). During that interval the training group received training, but the control group did not. The training group also received a follow-up posttest one month later (Time 3).

Participants

The participants included 135 native speakers of Chinese—in middle childhood, early adolescence, and young adulthood—who had all been learning English as an L2 since kindergarten or first grade (around age 5 or 6). They included 61 4th graders (aged 9 to 10 years), 38 7th and 8th graders (aged 11 to 13 years), and 36 university students (aged 18 to 22

years) in Hong Kong. Among the university students, 26 named Cantonese as their first language and 9 named Mandarin; one did not provide the information. All the young participants (in 4th to 8th grade) were native speakers of Cantonese. Almost all of their English teachers had spoken Chinese as L1 and English as L2. Except for six of the university students who had participated in exchange programs overseas for one semester, the participants had not lived in any English-speaking country. Hence, their exposure to native speakers of English had been very limited.

The training group consisted of 29 4th graders (48% boys), 23 7th/8th graders (48% boys), and 18 university students (33% men); for the control, there were 32 (50% boys), 15 (20% boys), and 18 (28% men) students respectively in each age group. Prior to data collection, we obtained written informed consent from the adults and written parental consent for the children, as well as verbal assent from the children themselves. All procedures performed in this study involving human participants were in accordance with the ethical standards of the Human Research Ethics Committee at the first author's institution.

Stimulus materials

Perceptual training program. We modeled our training program on other HVPT studies that involved both children and adults (Giannakopoulou et al., 2013, 2017); we used four native speakers of American English to present identical input to native Chinese speakers. We then compared how well the children and adults in our study learned the contrast in English between /p, t, k/ and /b, d, g/, after years of probably using aspiration rather than voicing or preceding-vowel length to distinguish these stops in word-medial or word-final position.

During the 20-session training, participants logged on to a computer training program in which they were entreated by cartoon characters to help in a range of scenarios (e.g., to rescue

people from a sinking ship or volcanic eruption, to help find a lost friend or lost pet). Participants earned different tools (e.g., lifeboats and lifesavers in the sinking ship scenario) to accomplish the mission at each level of the game by completing the trials in the training sessions. They were promoted to the next level when sufficient trials had been completed, and the number of tools collected was directly related to their performance on the trials. Screenshots of a training session are presented as supplementary material online.

The self-paced sessions were presented on laptop computers using E-Prime 2.0 software and portable headphones (Audio Technica ATH-ON303). Each session lasted about five minutes and contained 72 trials. Here is an example of a trial: Learners heard the audio input ("I say stable"), saw the words "*staple"* and *"stable"* on the screen, indicated by a keyboard press which word they had just heard, and then received feedback (a jingle for correct response or a buzz for incorrect response). At the end of each session, the computer showed the percentage of correct answers for that session and how many "tools" had been earned. The 4^{th} and $7^{th}/8^{th}$ graders in this study were rewarded according to the total number of tools they collected over the entire training program (i.e., bigger prizes for better performance during training). The rewards (e.g., snacks, stationery items) were modest, but our pilot study suggested that these reinforcers helped to motivate the learners.

There were altogether 864 training tokens: 4 native speakers of American English (3 females and 1 male; students from UCLA in their early 20s) each recorded 3 times of 72 training phrases (specifically, 6 target words for each of the 6 consonants /b, p, d, t, g, k/ in 2 positions, namely word-medial and word-final, but not for word-initial; see Appendix). All 864 tokens were used at least once and no more than twice for each participant in the entire training program. In each of the 20 training sessions, 72 training tokens were randomly drawn with the

constraints that they were evenly divided among the 6 target consonants and the 2 within-word positions. Similar to Giannakopoulou et al. (2013, 2017), we employed a within-session talker variability design in the training, for which tokens recorded by the 4 speakers were randomly presented within and across training sessions. Other studies reported using comparable number of speakers—e.g., 5 speakers in Lively et al. (1993) and Bradlow et al. (1997, 1999), but participants in those studies heard only one talker per training session. Implications arising from the distinction in talker variability will be further discussed in a later section.

We pilot tested the words used in the training phrases on 96 2nd graders in Hong Kong by asking them to read aloud those words. The majority of the 2nd graders were able to read out most of the words used in the training. This ensured that the youngest participants in our study (i.e., the 4th graders) were likely to recognize those words.

Assessment. The perception assessment included 150 trials and took about seven minutes to complete. The assessment trials resembled the training trials but without corrective feedback. Seventy-two of the test trials used the training words, while the other 78 test trials used words not in the training word set in order to measure how well the training generalized (Appendix). All test phrases were recorded by a native speaker of American English (also a student from UCLA) who had not been involved in recording the training phrases.

Procedure

Participants in each age group were randomly assigned to either the training or control group. They first took a pretest at Time 1. Participants in the training group then received self-paced perceptual training—20 five-minute sessions over 4 to 6 weeks. Training sessions were conducted 3-5 times per week at the participants' schools for the 4th and 7th/8th graders, and at the second author's laboratory for the university students. Within one week after this training

program, participants took a posttest (immediate posttest; Time 2), and they took another posttest a month later (follow-up posttest; Time 3). Participants in the control group had two baseline assessments to match the pretest (Time 1) and immediate posttest (Time 2) of the training group. We explained that multiple testing during the waiting period assessed the stability of their speech perception across time, and they received the training program after the Time 2 assessment.

The randomized controlled design enabled us to evaluate the effects of the training program by comparing the changes from Time 1 to Time 2 in the training group versus in the control group. The 1-month posttest on the training group (Time 3) allowed us to see whether training benefits detected at the immediate posttest (Time 2), if any, were maintained for at least one month. Besides the rewards that participants received for correct responses as mentioned earlier, adults received either HKD$30 (about USD$4) or a half-hour research participation credit for a psychology course for each assessment session (three for the training group and two for the control group). They also received a completion bonus of HKD$200 (about USD$25) for doing all the assessment sessions. The younger participants received a completion bonus too, in the form of small gifts.

Results

Table 1 presents the mean correct percentages (with standard deviations) for the perception task for each age group at Time 1 and Time 2.

<INSERT TABLE 1 ABOUT HERE>

H1: HVPT with immediate corrective feedback improves Chinese speakers' perception of stop consonants in L2 English

We examined the training effects on the perception accuracy using ANCOVA, with Time 2 overall mean correct percentage entered as the dependent variable, the experimental condition (training vs. control) and age group (children, adolescents, adults) as the between-subject factors, and Time 1 overall mean correct percentage as the covariate.

The main effects of experimental condition ($F[1,128] = 24.18$, $p < .001$, $\eta_p^2 = .16$) and age group ($F[2,128] = 17.46$, $p < .001$, $\eta_p^2 = .21$) were both significant, while condition X age group interaction was not significant ($F[2,128] = 1.45$, $p = .24$, $\eta_p^2 = .02$). Specifically, the training group significantly outperformed the control group at Time 2, after controlling for their Time 1 overall accuracy (mean difference = 4.52, SE = .92, $p < .001$, Cohen's $d = .84$). Hence, H1 was supported.

We conducted a *minF'* analysis to see if this training effect would generalize to untrained words (Raaijmakers, 2003; Raaijmakers, Schrijnemakers, & Gremmen, 1999). The *minF'* statistic was computed from F_1 and F_2 based on the equation $minF' = (F_1 \times F_2) / (F_1 + F_2)$, where F_1 and F_2 were the F-ratios for the treatment effect in subject analysis and item analysis respectively (Raaijmakers, 2003). For subject analysis, we performed a repeated-measure ANOVA on the change scores (Time 2 score minus Time 1 score) with the two word categories (trained vs. untrained) as the within-subject factor, and age and condition as the between-subject factors. The main effect of word category was not significant ($F_1[1,130] = 2.19$, $p = .14$, $\eta_p^2 = .02$). In item analysis, data were collapsed across subjects, and the effect of word category was not significant ($F_2[1,148] = 2.19$, $p = .14$, $\eta_p^2 = .02$). The composite measure *minF'* was not significant for trained versus untrained word categories ($minF'[1,277] = 1.10$, $p = .30$), suggesting that the training effects on word perception held across trained and untrained words (Clark, 1973).

H2: HVPT benefits both adults and children

According to the ANCOVA conducted earlier to test H1, the training effects on the perception of stop consonants did not vary significantly across age groups, as indicated by a non-significant condition X age group interaction (F[2,128] = 1.45, p = .24, η_p^2 = .02). Post hoc pairwise comparisons with Bonferroni adjustment showed significant differences in performance between the training and control groups at Time 2, favoring the training condition for all age groups after controlling for Time 1 score (children: mean difference = 2.93, SE = 1.33, p = .03, Cohen's d = .54; adolescents: mean difference = 3.94, SE = 1.71, p = .02, Cohen's d = .75; adults: mean difference = 6.70, SE = 1.76, p < .001, Cohen's d = 1.09). Hence, HVPT was found to benefit both adults and children, supporting H2. The results are illustrated in Figure 1.
<INSERT FIGURE 1 ABOUT HERE>

H3: Training benefits of HVPT will maintain for at least one month after training

Results just reported already demonstrated HVPT training benefits at immediate posttest using the randomized controlled experiment design. Repeated measures ANOVAs were then conducted separately on each age group for the training condition to see if the training benefits detected at immediate posttest (Time 2) could maintain for at least one month until Time 3 (Table 2). For each age group, time (Time 1, Time 2, Time 3) was entered as a within-subject factor to check the maintenance effect.

The results are presented in Table 2 and Figure 2. The main effect of time was significant for both the children (F[2,56] = 7.96, p = .001, η_p^2 = .22) and the adults (F[2,32] = 20.62, p < .001, η_p^2 = .56), but not for the adolescents (F[2,44] = 2.30, p = .11, η_p^2 = .10). Word perception

at Time 3 for the adolescents was not significantly different from either Time 1 (mean difference = 1.22, SE = 1.46, p = 1.00, Cohen's d = .19) or Time 2 (mean difference = -1.54, SE = 1.05, p = .47, Cohen's d = .20). For both children and adults, word perception at Time 3 was significantly better than Time 1 (children: mean difference = 4.12, SE = 1.19, p = .005, Cohen's d = .68; adults: mean difference = 7.69, SE = 1.60, p = .001, Cohen's d = .98) but not than Time 2 (children: mean difference = .64, SE = .83, p = 1.00, Cohen's d = .09; adults: mean difference = .12, SE = .59, p = 1.00, Cohen's d = .02). Hence, the training benefits for these two age groups were maintained for at least one month, thereby supporting H3.

<INSERT TABLE 2 ABOUT HERE>

<INSERT FIGURE 2 ABOUT HERE>

H4: HVPT improves perception of English stop consonants in the word-medial and word-final positions, but not in the word-initial position

Table 1 shows the mean correct percentages for the perception of stop consonants in different word positions. Three separate ANCOVAs were perfomed on the perception of word-initial, word-medial, and word-final consonants respectively. In each ANCOVA, Time 2 perception was the dependent variable, with experimental condition and age group as the between-subject factors, and Time 1 score as the covariate. The results are illustrated in Figure 3.

As predicted by H4, the main effect of experimental condition was significant for the word-medial (F[1,128] = 11.62, p = .001, η_p^2 = .08) and word-final positions (F[1,128] = 34.86, p < .001, η_p^2 = .21), but not for the perception of word-initial stop consonants (F[1,128] = .42, p = .52, η_p^2 = .003) for which Chinese speakers should be good at prior to the training because the contrast between /p, t, k/ and /b, d, g/ in word-initial position was the familiar aspiration also

found in their L1 Chinese. In fact, performance was generally high for word-initial stops in all age groups at Time 1, with the mean perception accuracy above 90% for the children and adolescents and above 97% for the adults (Table 1). This provided support to our assumption that Chinese speakers were already good at distinguishing the word-initial stops in English prior to the training.

For the word-medial position, condition X age group interaction was significant (F[2,128] = 3.21, p = .04, η_p^2 = .05). Posthoc pairwise comparisons with Bonferroni adjustment showed that HVPT significantly improved Time 2 perception of word-medial consonants in the adult group (mean difference = 8.40, SE = 2.28, p < .001, Cohen's d = 1.17), but not among the children (mean difference = 1.27, SE = 1.71, p = .46, Cohen's d = .19) and adolescents (mean difference = 2.55, SE = 2.21, p = .25, Cohen's d = .38).

For the word-final position, condition X age group interaction was not significant (F[2,128] = .31, p = .74, η_p^2 = .005). Posthoc comparisons indicated significantly better performance at Time 2 for the training condition relative to the control in all age groups (children: mean difference = 7.64, SE = 2.02, p < .001, Cohen's d = .92; adolescents: mean difference = 7.22, SE = 2.60, p = .006, Cohen's d = .90; adults: mean difference = 9.89, SE = 2.65, p < .001, Cohen's d = 1.05).

<INSERT FIGURE 3 ABOUT HERE>

Discussion

It is notoriously difficult for Chinese speakers to master the contrast between the English /b, d, g/ and /p, t, k/, which are ubiquitous in English. Nevertheless in this experiment, audio-recorded input from native English speakers, coupled with immediate corrective feedback, helped Chinese speakers learn to perceive distinctions among these six stop consonants. The

benefits of HVPT for word perception generalized across trained and untrained words, remained robust for at least one month, and were consistent with a priori theoretical predictions regarding within-word positions. Most significantly, our results documented that such training could help native Chinese speakers of several ages—children, early adolescents, and young adults—learn the challenging contrast between the ubiquitous voiced versus voiceless stop consonants in English.

With L2 education becoming more commonplace than ever, educational resources are severely stretched. When the L2 is not the societal language, native-speaker teachers are often in short supply and can be costly. While HVPT has been shown in prior research to benefit adult L2 learners, very few studies thus far have examined the effects of HVPT among children (Giannakopoulou et al., 2013, 2017; Shinohara, 2014; Shinohara & Iverson, 2013), and even fewer have compared the training effects across children and adults (Giannakopoulou et al., 2013, 2017; Shinohara, 2014). Our results generally corroborate the findings of prior studies conducted among other L2 learners of English, indicating that HVPT could benefit both children and adults (H1 and H2). Previous evidence of perceptual training benefits for adults learning L2 phonology (Bradlow et al., 1997, 1999; Iverson et al., 2012; McCandliss et al., 2002; Thomson, 2012) is thus extended to Chinese children as well as adolescents learning the challenging contrast between voiced and voiceless stop consonants in English. Therefore, audio recordings apparently can help school children acquire L2 phonology at different levels—at the individual speech sounds (i.e., phonemic) level as demonstrated in this study, and also at the global accent level (Au, 2013; Au et al., 2015).

Past studies on HVPT comparing children and adults have shown mixed results—some observed greater improvements in children than in adults (Giannakopoulou et al., 2013;

Shinohara, 2014), while others did not (Giannakopoulou et al., 2017). Our results showed that the overall training effects on the perception of English stop consonants did not vary significantly across age groups (H2). Hence, our study did not find evidence to support the notion of maturation constraints in L2 learning. The improvements in the adult group suggested that the degree of neuroplasticity in adults is still sufficient to enable perceptual adaptation.

Interestingly, adults in our training group appeared to reap more benefits from the training than the younger age groups for contrast of stop consonants in certain word position. Specifically, significant improvements were observed for both word-medial and word-final positions in the adult group, whereas the children and adolescents only showed improvement for the word-final consonants (H4). These results might suggest that the contrasts in word-medial intervocalic position, cued mainly by voicing, are more difficult for children to grasp than contrasts in word-final position, that are cued by both voicing and vowel duration (and at times aspiration).

Perceptual training presumably enhances speech perception by reallocating the learners' attention towards dimensions relevant to the classification of phonetic contrasts and away from dimensions that are irrelevant, to facilitate better mapping of auditory properties onto phonetic categories (Francis & Nusbaum, 2002; Iverson & Kuhl, 1995). High-variability training involving trial-by-trial talker variability necessitates the shifting of listeners' attention between cues across talkers and phonetic contexts, as different talkers might produce different relative weightings of acoustic cues for a particular contrast (Francis & Nusbaum, 2002). While it may allow more exemplars for the discrimination of relevant and irrelevant cues, this process of talker normalization likely demands substantial cognitive resources (Nusbaum & Magnuson, 1997). Indeed, prior studies have reported greater benefits from high-variability input for high aptitude

participants, whereas those of low aptitude benefited more from low-variability training (Antoniou & Wong, 2015; Peracchione, Lee, Ha, & Wong, 2007; Sadakata & McQueen, 2014). Giannakopoulou et al. (2017) also reported greater perceptual improvements in children after low-variability (single talker) training compared to high-variability (four talkers) training. We postulated that the word-final contrasts in this study might be more salient to the L2 learners than the word-medial consonants, such that the acquisition of acoustic cues to word-final stops was robust even for the children who presumably had less cognitive resources. Hence, our finding seemed to lend support to the hypothesis that the high-variability input induces additional processing costs among learners, and consequently produces less training benefits for younger participants who may have less cognitive resources for tackling the task (Antoniou & Wong, 2015; Barcroft & Sommers, 2005; Peracchione et al., 2007).

Note that while all the children and adolescents in this study were native speakers of Cantonese, about 25% of the young adults (i.e., university students) were native speakers of Mandarin. As such, the adult group was not entirely analogous to the homogeneous Cantonese-speaking younger age groups, and thus the results need to be interpreted with caution. Nonetheless, it is worth mentioning that word-final contrasts in English are expected to be more difficult to acquire by Mandarin speakers than Cantonese speakers, because Mandarin does not allow a stop as a coda consonant while Cantonese allows /p, t, k/ as well as a nasal as a coda consonant (Flege & Wang, 1989). Hence, the effect of including Mandarin speakers in the adult group, if any, should work against the observed training effect advantage of the adult group over the younger age groups. Therefore, we posit that our finding regarding the adult-over-children training advantage should still stand with homogeneity of Chinese dialects (i.e., all Cantonese speakers) among the age groups.

We also observed that the overall improvement in the perception of stop consonants was maintained for at least one month in both the children and adults, but not for the adolescents (H3). In fact, post-test performance was comparable between the two younger age groups, and the children even seemed to outperform the adolescents on the perception of word-final consonants at Time 2. We speculated that the relatively less robust improvement and long-term retention of learning in the adolescents might be partly attributed to the inadequacy of the training paradigm to fully engage the adolescents in this study. While we attempted to incorporate game-like features in the computer training program to make it more appealing to the younger participants, the attention-maintaining mechanisms embedded in the program might have worked better for the children than the adolescents who might not find this "game" as interesting and engaging as the real video games they typically played. The younger children—likely with less exposure to video games (Lee & Busiol, 2016)—might possibly be more motivated in the training. These speculations are yet to be explored and could matter in designing training programs for adolescents.

Our findings have important implications for L2 education. For instance, it has been suggested that adult L2 learners may be reluctant to seek face-to-face language input from native speakers for fear of being stigmatized or teased due to their foreign accents and grammatical errors (Au et al., 2017; Derwing & Rossiter, 2002; Gardner, 1979; Goto, Gee, & Takeuchi, 2002; Lee & Rice, 2007). Although the training program was perceptual in nature, prior research has suggested that such HVPT can also help adults improve their production of challenging speech sounds (Bradlow et al., 1997, 1999). Based on the current findings, modern technology can offer good L2 input via audio and video recordings to help improve L2 phonology among adult learners. We postulate that improving perception might in turn improve production probably

because accurate speech perception offers language learners a good internal model to guide them to modify their speech production to emulate native speakers (Au, Knightly, Jun, & Oh, 2002; Best, 1994; Bradlow et al., 1997; Flege, 1995). Downstream, the improvement in word perception documented in the present experiment might lead to improvement in the learners' word production as well, and thus help build adult L2 learners' confidence in seeking out and enjoying live interaction with native speakers. The effects of HVPT on enhancing L2 learners' production of English stop consonants await further investigations.

To conclude, the perceptual training benefits documented in our study worked well for children and adults, were robust across time, and generalized well from trained words to untrained ones, for the perception of basic ubiquitous sounds known to be challenging to learners of a second language. Such cost-effective training regimen can serve as a valuable model that could be incorporated into L2 education for learners of a wide age range.

Acknowledgments

We thank the participants, parents, and schools for their support of this research. We are grateful to our teams of dedicated lab assistants in Hong Kong and California for preparing the research materials and collecting and processing the data, as well as to Karen Ravn and Karin Stromswold for their feedback on earlier drafts. The research was supported by the Research Grants Council of Hong Kong (HKU740511H).

References

Antoniou, M., & Wong, P. C. (2015). Poor phonetic perceivers are affected by cognitive load when resolving talker variability. *The Journal of the Acoustical Society of America, 138*(2), 571-574.

Au, T. K. F. (2013). Songs as ambient language input in phonology acquisition. *Language Learning and Development, 9*(3), 266-277.

Au, T. K. F., Chan, W. W., Cheng, L., Siegel, L. S., & Tso, R. V. Y. (2015). Can non-interactive language input benefit young second-language learners?. *Journal of Child Language, 42*(2), 323-350.

Au, T. K. F., Knightly, L. M., Jun, S. A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science, 13*, 238-243.

Au, T. K. F., Kwok, A. F. P., Tong, L. C. P., Cheng, L., Tse, H. M. Y., & Jun, S. A. (2017). The Social Costs in Communication Hiccups Between Native and Nonnative Speakers. *Journal of Cross-Cultural Psychology, 48*(3), 369-383.

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition, 27*(3), 387-414.

Best, C.T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J.C. Goodman & H.C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*(5), 977-985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299-2310.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica, 22*(3), 129-159.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*(4), 335-359.

Conrad, A. W., & Rubal-Lopez, A. (Eds.). (2011). *Post-imperial English: Status change in former British and American colonies, 1940-1990* (Vol. 72). New York: Walter de Gruyter.

Curtiss, S. (2014). *Genie: A psycholinguistic study of a modern-day wild child*. New York: Academic Press.

Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System, 30*(2), 155-166.

Deterding, D., & Nolan, F. (2007, August). Aspiration and voicing of Chinese and English plosives. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 385-388). Universität des Saarlandes Saarbrücken Germany.

Flege, J. E. (1988). The development of skill in producing word-final English stops: Kinematic parameters. *The Journal of the Acoustical Society of America, 84*(5), 1639-1652.

Flege, J.E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-272). Baltimore: York Press.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1996). Factors affecting the production of word-initial consonants in a second language. *Second Language Acquisition and Linguistic Variation, 10*, 47-73.

Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/–/d/ contrast by native speakers of English, Mandarin, and Spanish. *The Journal of the Acoustical Society of America, 92*(1), 128-143.

Flege, J. E., & Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics, 17*, 299-315.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78-104.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 349–366.

Gardner, R. C. (1979). Social psychological aspects of second language acquisition. In H. Giles & R. St. Clair (Eds.), *Language and social psychology* (pp. 193–220). Oxford, UK: Basil Blackwell.

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ, 5*, e3209.

Giannakopoulou, A., Uther, M., & Ylinen, S. (2013). Enhanced plasticity in spoken language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English. *Child Language Teaching and Therapy, 29*(2), 201-218.

Goto, S. G., Gee, G. C., & Takeuchi, D. T. (2002). Strangers still? The experience of discrimination among Chinese Americans. *Journal of Community Psychology, 30*, 211–224.

Iverson, P., & Kuhl, P. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97*, 553–562.

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics, 33*(1), 145-160.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.

Kahn, D. (2015). *Syllable-based generalizations in English phonology*. New York: Routledge.

Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics, 16*, 153-169.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831.

Lee, T. Y., & Busiol, D. (2016). A review of research on phone addiction amongst children and adolescents in Hong Kong. *International Journal of Child and Adolescent Health, 9*(4), 433-442.

Lee, J. J., & Rice, C. (2007). Welcome to America? International student perceptions of discrimination. *Higher Education, 53*, 381–409.

Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge, England: Cambridge University Press.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384-422.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3), 1242-1255.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience, 2*(2), 89-108.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*(1), 11-28.

Nusbaum, H. C., & Magnuson, J. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 109–132). San Diego, CA: Academic Press.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research, 5*(3), 261-283.

Peracchione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2007). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America, 130*, 461–472.

Raaijmakers, J. G. (2003). A further look at the "language-as-fixed-effect fallacy". *Canadian Journal of Experimental Psychology*, *57*(3), 141-151.

Raaijmakers, J. G., Schrijnemakers, J. M., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language*, *41*(3), 416-426.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *The Journal of the Acoustical Society of America, 51*(4B), 1296-1303.

Sadakata, M., & McQueen, J. (2014). Individual aptitude in Mandarin lexical tone learning predicts effectiveness of high-variability training. *Frontiers in Psychology, 5*, 13–18.

Shinohara, Y. (2014). Perceptual training of English /r/ and /l/ for Japanese adults, adolescents and children. *Doctoral thesis*. United Kingdom: University College London.

Shinohara, Y., & Iverson, P. (2013, June). Computer-based English /r/-/l/ perceptual training for Japanese children. In *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, No. 1, p. 060049). ASA.

Shinohara Y, & Iverson P. (2015). Effects of English /r/-/l/ perceptual training on Japanese children's production. In The Scottish Consortium for ICPhS 2015, ed. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow.

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning, 62*(4), 1231-1258.

Walsh, T., & Parker, F. (1981). Vowel length and voicing in a following consonant. *Journal of Phonetics, 9*(3), 305-308.

Appendix

Training and testing words in "I say ___" phrases.

Pair: /p/ - /b/

Word-initial		***Word-medial***		***Word-final***	
pack*	back*	crappy*	crabby*	cap*	cab*
pay*	bay*	gapping*	gabbing*	lap*	lab*
peak*	beak*	napping*	nabbing*	nip*	nib*
pet*	bet*	rapid*	rabid*	rope*	robe*
pill*	bill*	bopping	bobbing	cop	cob
		mopping	mobbing	cup	cub
		nipple	nibble	mop	mob
		sopping	sobbing	nap	nab
		staple	stable	rip	rib
		swapping	swabbing	tap	tab

Pair: /t/ - /d/

Word-initial		***Word-medial***		***Word-final***	
teal*	deal*	atom*	Adam*	bet*	bed*
tie*	die*	blunter*	blunder*	bit*	bid*
time*	dime*	coating*	coding*	fat*	fad*
ton*	done*	patting*	padding*	fate*	fade*
tuck*	duck*	betting	bedding	bat	bad
		butting	budding	coat	code
		hinter	hinder	feet	feed
		metal	medal	got	god
		petal	pedal	mat	mad
		rating	raiding	not	nod

Pair: /k/ - /g/

Word-initial		***Word-medial***		***Word-final***	
cane*	gain*	backer*	bagger*	dock*	dog*
cap*	gap*	locking*	logging*	duck*	dug*
coat*	goat*	mucking*	mugging*	jock*	jog*
con*	gone*	plucking*	plugging*	peck*	peg*
cot*	got*	bicker	bigger	back	bag
		blocking	blogging	lock	log
		lacking	lagging	muck	mug
		locker	logger	pick	pig
		stacker	stagger	rack	rag
		tinkle	tingle	tack	tag

* = untrained words