Mandarin-English bilinguals' perception of self-produced non-native speech sounds

Madeleine Yu

Advisors: Dr. Sarah C. Creel (Cognitive Science), Dr. Marc Garellek (Linguistics), Reina Mizrahi (Cognitive Science)

Abstract

Language is an essential component of communication and much research has examined the trajectory of language acquisition throughout the course of development. However, the extent to which representations of native language speech sounds remain malleable is unknown. This current study examined the interactions between adult Mandarin-English bilinguals' native and non-native languages to determine the flexibility of native speech sound categories. Additionally, the influence of perception on production, and vice versa, was assessed. Through an eye-tracked identification task, participants of different language backgrounds (bilingual vs. monolingual) heard recordings of English words produced by two speakers. In the case of bilinguals, recordings featured self-produced non-native speech and the speech of a native speaker, while monolinguals heard recordings of both a non-native and a native other. Participants then completed a four-alternative-forced-choice task that involved selecting the picture that corresponded to each recording from four pictured options. A significant two-way interaction demonstrated that monolinguals performed with greater accuracy than bilinguals when hearing native speakers while bilinguals performed with greater accuracy than monolinguals when hearing their own non-native productions. This suggests the use of idiosyncratic perception and productions that enable non-native bilinguals to distinguish between ambiguous speech sounds and implies that categorical representations are, in fact, flexible in adulthood.

Background

The ability to communicate successfully with other individuals in the context of spoken language hinges upon the ability to both perceive and produce the distinctions between specific speech sounds. Evidence supports the concept of perceptual attunement, or the fine-tuning of native speech sound categories, which is thought to start early in the first year of life (Werker & Tees, 1984) with the onset of external linguistic input from the environment. This culminates in the ability for native speakers to perceive and produce language within the context of native speech categories (Rochet, 1995). Development of such representations is ongoing in children as old as five years old (Creel, 2017). While it remains unclear whether children's native word representations are based on the speech of frequently-heard speakers or on the child's own motor productions, evidence supports the precedence of native-like perception before native-like production (Cooper, Fecher & Johnson, 2017; Creel, 2017). Conversely, as speech sound categories within the first language (L1) become more stable, the ability to perceive differences between non-native speech sounds that are phonologically similar becomes more difficult (Kuhl et al., 1992). This contributes to the challenge of secondary language (L2) learning, such that for some non-native speakers of a language, productions are "accented" and therefore do not sound native-like. A particularly well-studied example of this is the difficulty which Japanese native speakers have when discriminating between the English /r/-/l/, a distinction which does not exist in Japanese (e.g., Goto, 1971; Miyazaki et al., 1975; Mochizuki, 1981; Best & Strange, 1992; MacKain et al., 1981). In children, early language acquisition is coupled with the development of motor skills and perceptual abilities, as well as a lack of prior language experience. All of which may play a part in children's successful improvement in L1 productions. Late learners of an L2, on the other hand, have the experience of skilled motor production in another language.

Researchers have examined the interactions between native and non-native languages during the acquisition of secondary language sound structure. There is indication of a "matched interlanguage speech intelligibility benefit (ISIB)" such that speakers and listeners who share the same language background identify foreign speech with equal or greater accuracy than speech between speakers and listeners who do not share the same language background (Bent & Bradlow, 2003). Such evidence suggests that L1 speech sound categories are not fixed, but are in fact flexible, such that non-native speech is still discernible between speakers who share the same language background. It is, however, unclear to what extent this benefit is dependent on either perception or production. Most examinations of the ISIB are between two individuals, a speaker and a listener, whose specific experience and distinct manner of using the non-native target language may differ (e.g., regional dialects). Little is known about the direct role of the individual's perceptual representations of non-native speech sounds on production and whether these representations are flexible or remain fixed within the L1 throughout life.

The Current Study

This project examines the direct intersection between perception and production by determining the extent to which non-native speakers of a language can understand themselves in comparison to hearing speech by a native L2 speaker. Specifically, Mandarin-English bilinguals whose first language is Mandarin and second language is English were tasked with listening to their own, and a native speaker's, productions of English words and identifying the corresponding pictured word from four presented options. Throughout this task, participant eye movements were tracked using the visual-world paradigm (VWP; Tanenhaus et al., 1995; Swingley et al., 1998). Each

English word to identify were paired with another that differs by one or two phonemic contrasts that are specifically distinguishable in English, but absent from Mandarin.

This study examines four competing hypotheses: first, it may be the case that non-native speakers demonstrate language development similar to that of children, such that their perception precedes production. While perception of L2 speech sounds may be native-like, productive abilities may lag and cause non-native speakers to understand native-produced speech better than their own**.** Second, it may be that representations of speech sounds are fixed. Thus, non-native speakers would only be able to perceive and produce the sounds that are distinct within their L1, Mandarin. Neither the minimal pairs produced by themselves nor by the native English speaker would reveal detectable differences. Third, perhaps non-native speakers are able to perceive and produce L2 speech sounds with near-to-native-like ability. Though they may speak accented English, native-like perception of the speech sounds could facilitate productions that are close enough to native representations. Prior research indicates the possibility of high comprehensibility in variable levels of accented speech (Munro & Derwing, 1995). Therefore, both the speech from the non-native self and the native other would yield detectable distinctions. Lastly, it may be the case that non-native speakers perceive and produce L2 speech sounds idiosyncratically. It is possible that the Mandarin-English bilinguals perceive and produce non-native speech neither wholly within the L1 or the L2, but in-between, resulting in differences that are uniquely distinct to individual non-native speakers. It may also be the case that the bilinguals share a language-general manner of producing English speech sounds and that use of the same strategies allows them to correctly identify the target word.

To examine whether non-native performance is indicative of the third or fourth hypothesis, an additional control group of native English monolingual listeners were recruited to perform the task. If it is the case that non-native speakers demonstrate native-like production, native listeners should perform with equally high accuracy when hearing speech from both speakers. However, if non-native perception and production is, in fact, idiosyncratic, native listeners should perform with less accuracy when hearing speech by a non-native speaker compared to speech by a native speaker.

Methods

Mandarin-English Bilinguals

Participants

46 out of a target population of 48 native Mandarin late-English bilinguals (32 Female), ages 18-24 (M = 19.5, SD = 1.23) were recruited through an online human participant pool and received academic credit. 13 additional participants were replaced in the cases where there was an undetectable accent (n = 3), less than 10% of looks to non-target pictures in at least one trial block (n = 2), greater than 15% difference in looks to the center of the screen across trial blocks (n = 3), equipment problems (n = 3), failure to meet the language background pre-requisites (n = 1), or experimenter error (n = 1).

Stimuli

A total of 32 words were used as the main stimuli of the experiment. Each item contrasted with a minimal pair by either a single or a combination of phonemic differences (3 /g/-/k/, 2 /d/-/t/ final voiced pairs, 3/m/-/n/ final pairs, 4 /ɪ/-/eɪ/, 1 /ɛ/-/æ/, 2 /ɛ/-/æ/ with /d/-/t/ final coda, and 1 /ɑ/-/ʌ/ with /g/-/k/ final coda) (e.g., *bag* – *back*, *card* – *cart*, *log* – *lock*). The phonemic differences were selected to be speech sounds that are not contrastive in Mandarin. Pictures selected to elicit each word were matched in size and participant familiarity. All of the pictures were standardized to 200x200 pixels each and extraneous information in the background of pictures was removed through GimpShop freeware.

Figure 1. Example of standardized picture stimuli

The auditory stimuli used were live recordings from each Mandarin-English bilingual speaker and one of 23 native-English speakers. Each speaker provided 32 natural productions by naming the pictured words when prompted by the researcher. For each participant, a total of 64 recordings were edited in PRAAT and excised of silences and other extraneous environmental sounds. All recordings were standardized to 70 dB.

The four pictures within each trial set remained constant throughout the study for all participants. Within each set, the pairs were matched to a phonetically and semantically different pair to increase clarity of the task.

Word Stimuli

Word	Pair	Phonemic Contrast	Type
Bag	Back	/g/-/k/	Coda
Bed	Bat	/ɛ/-/æ/, /d/-/t/	Vowel, Coda
Brake	Brick	/eɪ/-/ɪ/	Vowel
Cake	Kick	/eɪ/-/ɪ/	Vowel
Card	Cart	/d/-/t/	Coda
Chain	Chin	/eɪ/-/ɪ/	Vowel
Code	Coat	/d/-/t/	Coda
Comb	Cone	/m/-/n/	Coda
Dog	Duck	/ɑ/-/ʌ/, /g/-/k/	Vowel, Coda
Eggs	X	/g/-/k/	Coda
Gum	Gun	/m/-/n/	Coda
Head	Hat	/ɛ/-/æ/, /d/-/t/	Vowel, Coda
Lime	Line	/m/-/n/	Coda
Log	Lock	/g/-/k/	Coda
Pen	Pan	/ɛ /-/æ/	Vowel
Shape	Ship	/eɪ/-/ɪ/	Vowel

Procedure

The experiment consisted of four phases: (1) familiarization phase, (2) production phase, (3) recognition phase, and (4) post-study language test.

During the familiarization phase, participants were tasked with rating their familiarity with the words paired with the pictures on a 6-point Likert scale. Each word and picture were presented one-at-a-time in the center of screen by a Macintosh Mini using PsychoPy software

(Peirce, 2007). This task was used to measure participants familiarity with the words in the experiment and disambiguate some picture labels (*ship vs. boat)*. Pictures were presented in randomized order across participants.

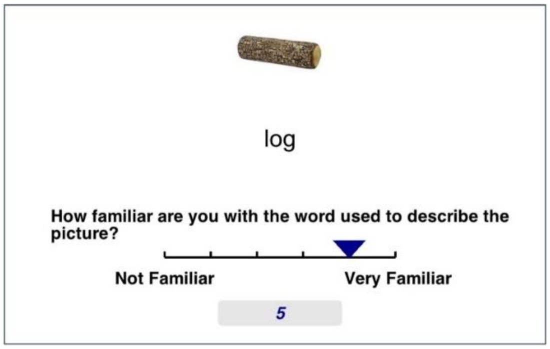

Figure 2. Example of familiarization task

Following the familiarization phase, the production phase began. Both a non-native bilingual speaker and a native monolingual speaker were recorded live. Pictures depicting each word appeared on the screen one after another and speakers were prompted by a researcher to name each aloud as they appeared. Audio recordings were made using the built-in microphone on Sennheiser GAME ONE headphones. All stimuli were presented using the Matlab Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997) and Eyelink Toolbox (Cornelissen, Peters & Palmer, 2002) on the MacOSX operating system. Participants were tested in a sound isolated room, and audio was presented using Sennheiser GAME ONE headphones with which participants could adjust loudness level.

At the beginning of this phase, non-native speakers were given five practice trials to familiarize them with the task and allow for any equipment adjustments to be made. The order of picture appearance was randomized for both speakers. Additionally, the order of which speaker was recorded first was counterbalanced across subjects so that exactly half of the non-native speakers were recorded first and the native speakers recorded second. Both speakers were present for all recordings to ensure the same exposure to environmental stimuli across participants. Words that were highly unfamiliar elicited a “don’t know”. All data for these words and their pairs were removed from analysis from both speaker recordings. Once both speakers were done recording, participants stepped out of the recording room and edits were made with PRAAT.

During the recognition phase, participants were tested in a four-alternative-forced-choice (4-AFC) task as their eye-movements were tracked using an Eyelink 1000 Remote eye tracker (SR Research, Mississauga, ON, Canada; http://www.sr-research.com). Each participant completed two blocks of 64 trials; one block featured the audio recordings by the non-native speaker and the other featured the recordings by the native speaker. Each trial included a set of four pictures of two minimal pairs (e.g., *lock-log-pen-pan)*. Picture sets were pseudorandomized for each participant so that no set reappeared within four trials and target position did not repeat more than twice in a row. In each trial, the audio stimuli played 500ms after the presentation of the trial set. Picture positions and the order of non-native and native recording blocks were counterbalanced across all participants.

Figure 3. Example of a trial set during the recognition phase

During the post-study language test, participant expressive vocabulary in English and Mandarin was measured using the Multilingual Naming Test (MiNT; Gollan et al., 2012). During the vocabulary assessment, participant responses were recorded for possible later offline scoring by a native Mandarin-speaking research assistant. At the end of the experiment, participants completed an in-depth language background questionnaire which focused on identifying participants familiarity with different dialects of Mandarin.

Native English Speakers

Participants

26 native English speakers (15 Female). 20 English-Monolinguals were recruited through an online psychology pool and received academic credit. Six additional speakers volunteered on multiple occasions in the event that a Mandarin-English bilingual did not have an English monolingual speaker paired with them during the experiment session. Of the additional speakers, two were early learners of English whose speech was indistinguishable from that of native speakers, while the other four speakers were monolingual English research assistants.

Stimuli

The same visual stimuli in the Mandarin-English bilingual condition were used. Each native English speaker provided half of the total auditory stimuli during the Mandarin-English bilingual experiment sessions and recorded themselves for each of the 32 words.

Procedure

The procedure matched that of the Mandarin-English bilingual condition, however, native speakers did *not* complete the recognition phase. Instead, the speakers completed (1) familiarization phase, (2) production phase, and (3) post-study language test.

All three parts were identical to that of the Mandarin-English bilinguals.

English Monolingual Listeners

Participants

21 out of a target population of 48 English monolinguals (16 female), ages 18-29 (M = 21.45, SD = 2.67). Participants were recruited through an online human participant pool and received academic credit. Six additional participants were replaced in the cases where there was greater than 15% difference in looks to the center of the screen across trial blocks (n = 5) or recordings used were made by bilingual participants who were later replaced (n = 1).

Stimuli

The same visual stimuli in the Mandarin-English bilingual condition were used. The auditory stimuli consisted of re-using both the non-native speaker and native speaker recordings from the Mandarin-English bilingual experiment sessions.

Procedure

The procedure matched that of the Mandarin-English bilingual condition, however, English monolinguals did *not* complete the production phase. English monolinguals completed (1) familiarization phase, (2) recognition phase (re-using recordings from the Mandarin-English bilingual experiment sessions) and (3) post-study language test.

The familiarization phase was identical to that performed by the Mandarin-English bilinguals. All English monolingual participants, with exception to one, completed the familiarization phase. In the recognition phase, the procedure was identical to that of the Mandarin-English bilinguals; however, the auditory stimuli did not include recordings of the participant themselves, but of a previously recorded non-native and native speaker pair. The purpose of reusing the recordings is to assess monolingual perception of the same recordings heard by the non-native listeners.

During the post-study language test, monolingual participants completed the English MiNT and the in-depth language questionnaire. In addition, monolinguals were asked to self-report the relative accentedness of each speaker heard in the recognition phase and to estimate the number of native Mandarin speakers they interact with hourly, on a weekly basis.

Preliminary Results

Behavioral data

A mixed-effects logistic regression was conducted on test accuracy as a function of listener group language background (Mandarin-English bilingual, English monolingual) and speaker type condition (non-native English bilingual, native English). Bear in mind that data collection is still ongoing, and thus analyses should be taken as preliminary. Additionally, data from trials that featured words that were misnamed by participants were removed from analysis, along with that of each word's pair from both speakers. Each participant maintained more than 16 out of 32 usable recordings during the recognition phase. A generalized linear mixed-effects model was used to examine the interaction between the independent variables, with random slopes and intercepts for participants and items. There was a significant effect of listener group language background (*estimate* = 0.308, *SE* = 0.129, *p* = 0.0169) with Mandarin-English bilinguals' accuracy lower overall in comparison to English monolinguals. A significant effect of speaker type condition was also observed (*estimate* = -0.489, *SE* = 0.101, $p < 0.0001$) such that recordings by the native speaker were more distinguishable (94.1%) than recordings by the non-native speaker (85.2%) across both listener groups. Additionally, there was also a significant two-way interaction (*estimate* = -0.636, *SE* = 0.959, $p < 0.0001$) between listener groups and speaker type conditions. Using a planned pairwise comparison to examine the exact nature of the interaction, it was revealed that Mandarin-English bilinguals performed significantly better in the non-native speaker condition than English monolinguals (*estimate* = 0.70, *SE* = 0.256, *p* = 0.0057), who, in turn, performed significantly better in the native speaker condition than the bilinguals (*estimate* = -2.04, *SE* = 0.419, $p < 0.0001$).

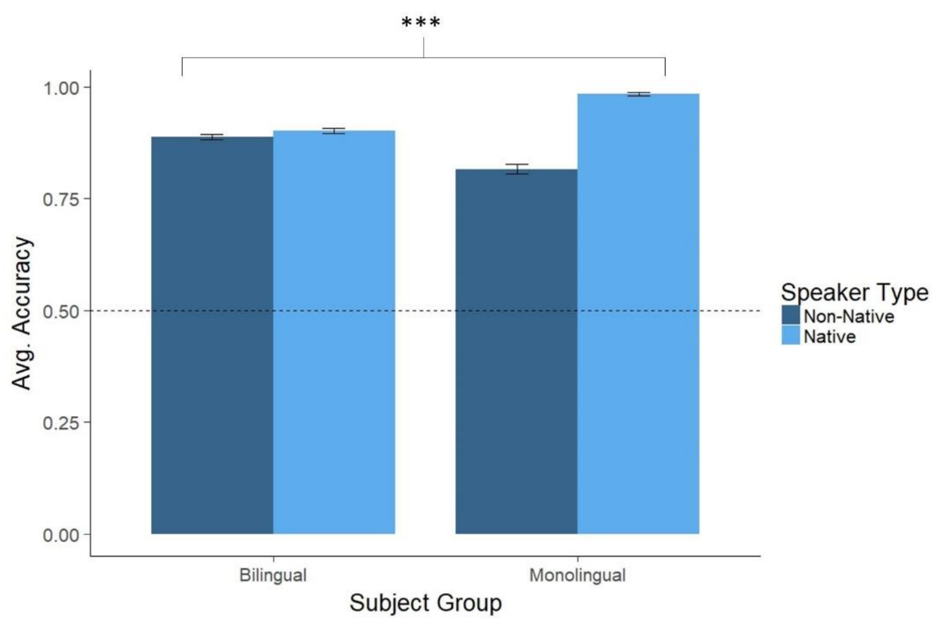

Figure 4. Average percent accuracy by listener group in each speaker type condition, ****p* < 0.0001

Eye-tracking data

Looks to the competitor picture were subtracted from looks to the target picture, to form a target preference looking score. Prior to subtraction, looking proportions underwent an empirical logit transformation to correct for non-normality (Barr, 2008). A specific time window of 200 to 1200 ms was used for analysis provided that it takes about 200 ms to plan and execute an eye movement and 1200 ms is typically the endpoint used in prior eye-tracked studies at the word level. Also, fixation data from trials that featured misnamed words by participants were removed from analysis, along with that of each word's pair from both speakers. Each participant maintained more than 16 out of 32 usable recordings for visual fixations during the recognition phase.

Time window (200-1200 ms). An analysis of variance (ANOVA) was conducted on the mean difference in empirical-logit-transformed look times to target picture minus competitor with listener group language background (Mandarin-English bilingual, English monolingual) and speaker type condition (non-native English bilingual, native English) as independent variables. We refer to this number as the target advantage. There was a significant effect of listener group language background ($F(1, 65) = 28.33$, $p < 0.0001$), where the monolingual participants showed greater target advantage than the bilingual participants overall. There was a significant effect of condition ($F(1, 65) = 33.37$, $p < 0.0001$), in which participants showed greater target advantage when hearing native speaker recordings rather than when hearing non-native speaker recordings. Similar to the behavioral data, there was also a significant effect of listener group language background on condition ($F(1, 1) = 48.78$, p < 0.0001). Monolingual listeners ($M = 1.99$) showed greater target advantage in the native speaker condition than bilingual listeners ($M = 1.01$) did ($t(47) = -9.84$, $p < 0.0001$). However, counter to the findings in the behavioral data, there was no indication of a target advantage for bilingual listeners ($M = 0.940$) in the non-native speaker condition compared to monolingual listeners (*M* = 0.944) ($t(35) = -0.02$, $p = 0.981$). This difference between the behavioral and eye tracking data may be due to the removal of fixation data from trials that featured misnamed word stimuli.

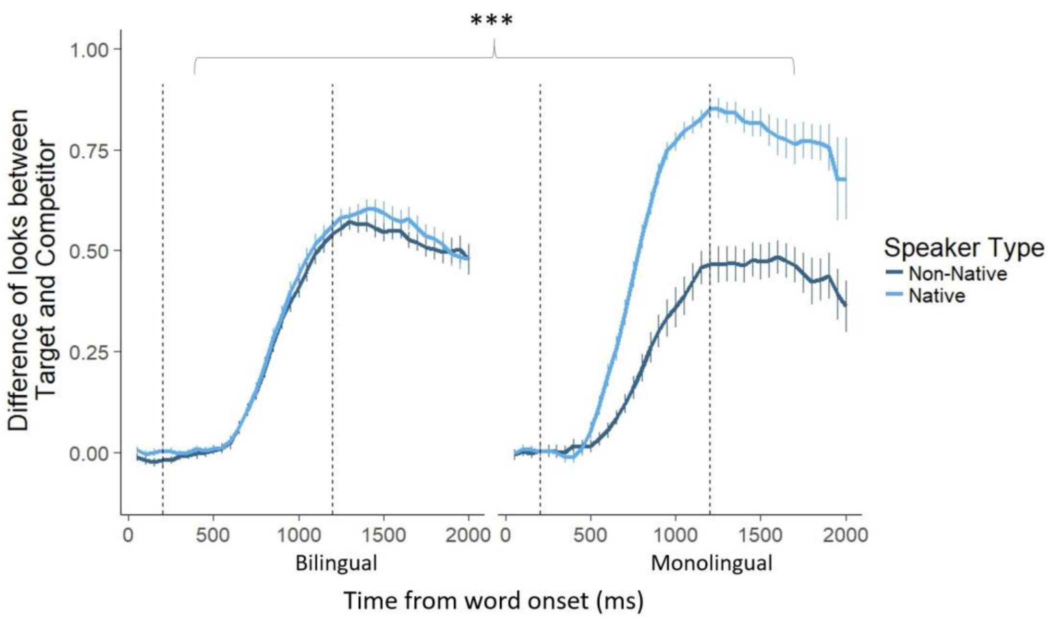

Figure 5. Looks to the target minus the competitor by listener group in each speaker type condition, ***$p < 0.0001$

Discussion

The findings from this preliminary analysis suggest, first and foremost, that the speech sound categories within Mandarin-English bilinguals are not fixed within their L1. Instead, bilinguals were above chance at distinguishing speech sound differences from both the native speaker and from their own non-native speech. Conversely, English monolinguals were also able to perceive distinctions in the non-native speech, though to a lesser extent than when they listened to speech by a native speaker. This indicates that regardless of the specific language experience or the age at which a secondary language is acquired, categorical representations of speech sounds remain malleable well into adulthood.

Another point of interest is the fact that bilingual non-natives performed with greater accuracy when hearing their own self-produced recordings in comparison to native English monolinguals hearing the non-native speech. This ability of bilinguals to better distinguish the differences in phonemically similar speech sounds within the non-native speech condition than monolinguals could, suggests that non-native speakers were using alternative, possibly idiosyncratic cues to perceive and produce distinctions. Given that there was a significant difference in the two listener groups' performances between the two speaker-type conditions, it demonstrates that the use of idiosyncratic cues was uniquely detectable to the Mandarin-English bilingual non-native participants alone. Regarding the previous idea that speech sound categories are malleable within both participant groups, this finding suggests that such idiosyncratic cues may develop from the specific interactions between the non-native bilinguals' two languages: Mandarin and English. In comparison, English monolinguals lack the experience of communicating in Mandarin and this possibly accounts for the inability to perceive the specific cues that bilinguals used to perceive and produce differences in the English speech sounds. While this seems to support the concept of the ISIB within self-produced non-native speech, it is important to note that it is still unclear whether the benefit arose from the bilinguals hearing their own speech and distinct manner of speaking, or if the use of idiosyncratic cues is a language-general effect that is shared across all Mandarin-English bilinguals.

With respect to the greater application of this study to language learning in general, it makes sense to ask whether learners of an additional language benefit more from hearing speech by a native speaker of the target language or if it is more helpful for them to learn a language from another non-native speaker who shares the same language background and experience. In this study, there is somewhat of a benefit for bilinguals in comparison to monolinguals for hearing English speech sounds in the context of non-native speech. Bilinguals demonstrated greater accuracy than monolinguals in the non-native speech condition; however, there is no apparent difference in bilinguals' performance between conditions. In other words, although bilinguals were better than monolinguals at correctly guessing what the intended target word was in the non-native speaker condition, their performance was no better in the native speaker condition. Thus, there appears to be no additional benefit between either direction for non-native listeners nor is there any noticeable detriment from hearing L2 speech sounds from a non-native speaker with the same language background.

In examination of the extent to which the bilinguals' perception influenced production, the findings of this study offer support in the direction that perception directly mediates production. Although there appears to be some perceptual limitations for the bilingual non-natives given that performance accuracy remained around 89%, the fact that they could perceive their own self-produced non-native speech just as well as the native-produced speech indicates that perceptions largely matched productions.

In conclusion, it appears that the representation of native speech sound categories, though structured within the boundaries of a distinct linguistic context, remains malleable even in adults. This flexibility allows for the opportunity to perceive and produce non-native speech sounds in a manner that, while perhaps is not quite native-like but is, rather, idiosyncratic, facilitates additional language learning beyond the confines of the first language.

Limitations and Future Directions

Although this study appears to support a language general benefit in which distinctions between non-native speech sounds are better comprehended by those who share the same language background than by those who do not, there are some limitations within this study that question the certainty of this finding. First of all, the bilingual non-native participants both recorded and then later heard their own productions in the experiment, while the monolingual group only heard the productions of a previously recorded speaker pair. It is therefore unclear whether the benefit demonstrated by the bilingual listener group in the non-native speaker condition was due to the fact that they heard speech that shared the context of their specific language background, or if it was entirely due to the familiarity of hearing their own speech. The next logical step to determine the cause of the benefit would be to recruit an additional group of Mandarin-English bilinguals to perform the same tasks completed by the English monolingual listener group in this study. If the benefit is simply due to hearing one's own manner of producing non-native speech sounds, the bilingual listeners may perform with a similar accuracy during the recognition phase to the monolingual listeners. However, presence of an interlanguage speech intelligibility benefit would otherwise maintain the current significant difference of speaker type condition between listener groups.

It also remains to be seen whether shared language general benefits, if any, are unique to the specific languages examined here. For instance, are there distinct effects on an individual's speech sound representations that depend on the particular secondary language acquired? In the case of this experiment, did the acquisition of English as a second language have specific consequences on how idiosyncratic perceptions and productions developed? Even in the context of this experiment, it is possible that some Mandarin-English bilinguals' productions within Mandarin was influenced by individual variability in dialectal experiences. If there is evidence of an interlanguage speech intelligibility benefit, the dialectal differences within Mandarin may be far too distinct to facilitate the benefit across individuals who ostensibly share the same language background. On the other hand, it could be the case that the benefit is facilitated through the particular experience of producing the target language in a non-native manner. It has been demonstrated that exposure to systematic variability in accented speech enables foreign-accent adaptation that is generalizable across speakers from differing language backgrounds (Baese-Berk, Bradlow & Wright, 2013). Perhaps the ISIB does not just pertain to individuals of the same language background, but enables non-native speakers with different linguistic experiences to better understand each other.

This current study utilized stimuli at the word level to test participants' ability to perceive and produce English speech sounds. Of course in a more naturalistic setting, individuals have the advantage of additional contextual information such as context, sentences, and gestures to parse out useful cues from ambiguity. While it is important to understand how each factor, especially at the level of individual words, is implicated within the greater scale of communication, it could be insightful to embed the words within semantically neutral sentences to examine how prolonged exposure to the non-native and native speaker productions effect listener performance. Perhaps the addition of more information may facilitate accent adaptation that enables English monolinguals to improve performance within the non-native speaker condition, or it is possible that earlier identification of the "accented" speech may yield increased confusability and lower accuracy overall. During the study, the recognition phase involved two trial blocks, one that featured the non-native speaker's productions and one that featured the native speaker's productions, counterbalanced across all participants. Thus participants could expect to hear only one individual's recordings throughout the course of the trial block. It may be of importance to compare the influence of trial block condition on performance by including an additional trial block that features intermixed recordings by the native and non-native speakers.

Lastly, an essential next step is to examine the specific acoustic properties of the Mandarin-English bilinguals' productions in comparison to that of the productions by native English speakers. It is possible that the differences in the speaker pairs' productions is a main factor in listener performances. English speech sounds that feature voicing differences in stop consonants, specifically preceding vowel duration, final consonant closure duration, and voicing in closure for final consonants, has been identified as a main source of misidentification for Mandarin speakers (Xie & Fowler, 2013; Bent & Bradlow, 2008; Hayes-Harb et al., 2008). Clarification of the exact properties within English speech sounds and how productions differ across speakers of variable experience communicating within the English language is crucial to disentangling the specific features that are at play in the intersection of perception and production within non-native L2 learners.

References

Baese-Berk, M.M., Bradlow, A.R., & Wright, B.A. (2013) Accent-independent adaptation to foreign accented speech. Journal of the Acoustical Society of America, 133(3), EL174-EL180. PMCID:P PMC3579861.

Barr, D. J. (2008) Analyzing 'visual world' eyetracking data using multilevel logistic regression. Journal of Memory and Language, 59, 457–474.

Bent, T., & Bradlow, A. (2003) The interlanguage speech intelligibility benefit. Acoustical Society of America.

Bent, T., & Bradlow, A.R. (2008) Perceptual adaptation to non-native speech. Cognition. 106, 707-729.

Best C.T., & Strange W. (1992) Effects of phonological and phonetic factors on cross-language perception on approximants. J Phonetics. 20:305–330.

Brainard, D. H. (1997) The Psychophysics Toolbox. Spatial Vision, 10, 433–436.

Cooper, A., Fecher, N., & Johnson, E.K. (2018) Toddlers' comprehension of adult and child talkers: Adult targets versus vocal tract similarity. Cognition, 173, 16-20

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002) The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. Behavior Research Methods, Instruments, and Computers, 34, 613–617.

Creel, S. C. (2017, November) What am I saying? Exploring the perception-production loop in children's speech. Talk presented at the Auditory Perception, Cognition, and Action meeting, Vancouver, BC, CA.

Gollan, T., Weissberger, G.H., Runnqvist, E., Montoya, R.I., Cera, C.M. (2012) Self-ratings of Spoken Language Dominance: A Multi-Lingual Naming Test (MINT) and Preliminary Norms for Young and Aging Spanish-English Bilinguals. Biling (Camb Engl). Jul; 15(3):594-615. PMID: 25364296.

Goto, H. (1971) Auditory perception by normal Japanese adults of the sounds 'L' and 'R' Neuropsychologia. 9:317–323.

Hayes-Harb, R., B.L. Smith, T. Bent & A.R. Bradlow. (2008) The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. Journal of Phonetics 36, 4, 664-679. Published, 2008.

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.

MacKain, K.S., Best, C.T., & Strange, W. (1981) Categorical perception of English /r/ and /l/ by Japanese bilinguals. Applied Psycholinguistics. 2:369–390.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J., & Fujimura, O. (1975) An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Percept Psychophys. 18:331–340.

Mochizuki, M. (1981) The identification of /r/ and /l/ in natural and synthesized speech. J Phonetics. 9:283–303.

Munro, M.J. & Derwing, T.M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. Language Learning. 1:73-97

Peirce, J.W. (2007) PsychoPy - Psychophysics software in Python. J Neurosci Methods, 162(1-2):8-13

Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10, 437–442.

Rochet, B. (1995) Perception and production ofL2 speech sounds by adults. In W. Strange (Ed.) Speech Perception and Linguistic Experience: Theoretical and Methodological Issues (pp. 379-410). Timoniwn, MD: York Press.

Swingley, D., Pinto, J. P., & Fernald, A. (1998) Assessing the speed and accuracy of word recognition in infants. Advances in Infancy Research, 12, 257-277.

Tanenhaus, M.K., & Trueswell, J.C. (1995) Sentence comprehension. In J. Miller and P. Eimas, eds., Speech, Language, and Communication, 217-262. San Diego, CA: Academic Press.

Werker, J. F., & Tees, R. (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development, 7, 49–63.