29th August 2017

MA thesis

Distributional learning of extrinsic vowel duration differences by Mandarin native speakers

MA General Linguistics

Yaming Zhang (11316977)

Faculty of Humanities, University of Amsterdam

Supervisor: Paul Boersma

Abstract

This study is undertaken in order to examine the effect of distributional learning on the acquisition of extrinsic vowel duration differences by Mandarin native speakers, with a focus on voicing-conditioned vowel duration. The perception of syllable-final stop voicing is one of most common difficulties that Mandarin native speakers encounter during English learning. Mandarin speakers usually have difficulties in perceiving syllable-final consonant voicing when the final consonant is unreleased. This is considered to be due to negative transfer from their first language.

The specific research question was whether distributional learning on vowel duration differences is more effective when syllable-final voicing feature is also available during the training. In order to achieve the research goal, an experiment was conducted on 40 Mandarin native speakers.

The result revealed a significant difference between released bimodal group and released unimodal group ($p = 0.018$). At the same time, no significant triple interaction was found among distribution type (bimodal vs. unimodal), test type (pre-test vs. post-test) and stimulus type (released vs. unreleased). In other words, although there was a significant improvement in the released bimodal group, we cannot draw a conclusion that training is more effective with released stimuli. Further research needs to be done to confirm that the improvement in the released bimodal group is due to the factor of syllable-final voicing.

Key words: second language acquisition, syllable-final stop voicing, vowel duration, distributional learning

Table of Contents

1. Introduction

This study is undertaken in order to examine the effect of distributional learning on the acquisition of extrinsic vowel duration differences by Mandarin native speakers, with a focus on voicing-conditioned vowel duration.

In second language acquisition, one’s first language (L1) has potential impact on the acquisition process, and the influence of L1 could appear in both perception and production. When it comes to English learning, learners with different language backgrounds usually encounter different problems. For Mandarin speaking English learners, one of the difficulties is the perception of syllable-final stop voicing. It was reported by previous studies that Mandarin Chinese speakers are able to perceive released syllable-final voicing very well, but they show difficulties in perceiving unreleased syllable-final stop voicing (Flege 1989; Zhang & Hayashi 2015). For instance, Mandarin speakers can only distinguish the minimal pair /bed/ and /bet/ if they hear the full pronunciation. When they hear [mat̚_] and [ma:d̚_], they won’t be able to tell the words apart. What causes this difficulty?

First, a voicing contrast of stop consonants in syllable-final position is unfamiliar to Mandarin native speakers. It has been long recognized that the syllable is the basic unit in the perception of spoken language. The difference of syllable structure between L1 and L2 can potentially influence the acquisition of L2 sounds.

In English, a word with a syllable structure of CVC is very common, and almost every consonant (except for /h/, /w/ and /j/) can serve as a syllable-final consonant. Therefore, a voicing contrast of stop consonants occurs both syllable initially and syllable finally in English. However, Mandarin has a relatively simple syllable structure. There are in total 22 consonants and 14 vowels in Mandarin. All consonants except for /ŋ/ can appear in syllable-initial position, while only the nasal sounds /n/ and /ŋ/ are permitted to be syllable-final consonants. Stop consonants never occur in syllable-final position.

L2 learners often have difficulties in acquiring a contrast that occurs in an unfamiliar syllable position. In the case of unfamiliar syllable-final contrast, learners usually apply to syllable-final decisions the same procedures that they use for syllable-initial decisions.

Nevertheless, applying the same perceptual strategy could not deal with all pronunciation variants. In daily conversation, the reduction of sounds is very common, especially in syllable-final positions. For instance, the cue of stop release is often absent in American English.

Native speakers are usually more sensitive to secondary acoustic cues and they are able to perceive reduced pronunciation variants relying on limited cues. On the other hand, non-native speakers are thought to be less sensitive to certain secondary acoustic cues and they are only able to perceive fully pronounced words. The fact that Mandarin native speakers apply the same the perceptual strategy to syllable-initial contrast and syllable-final contrast makes it difficult for them to perceive those reduced pronunciation variants that are not fully released.

Flege (1989) investigated Mandarin Chinese speakers' ability to perceive English syllable-final /t/ and /d/ contrast. It is found that Mandarin speakers can perceive an unedited English /t/ and /d/ contrast in syllable-final position very well, but when the release burst and closure voicing part is removed, they performed much worse than English native speakers. The result indicated that Mandarin Chinese speakers tried to rely on syllable-initial cues to identify syllable-final contrast, and the released part in pronunciation is crucial for them in the perception of syllable-final stop voicing.

Zhang & Hayashi (2015) also examined Mandarin native speaker's perception of syllable-final consonants by using stimuli with a structure of $C_1V_1C_2C_3V_2$ with the consonant in C_2 position unreleased. In the study, the perceptual ability of Mandarin speakers was compared to that of Japanese speakers. It is found that Mandarin speakers had a worse performance compared to Japanese native speakers in distinguishing voiced and voiceless consonants in C_2 position.

Secondly, Mandarin native speakers are not sensitive to vowel duration difference. Vowel duration is one of the important acoustic features of vowels. In some languages vowel duration difference is phonemic, which means long and short vowel contrast can distinguish meaning. For instance, in Japanese /toke:/ means “clock” while /to:ke:/ means “statistics”. In some other languages, like English and Mandarin, vowel duration is not phonemic, which means that lengthening or shortening of a vowel does not change the meaning of a word. It is considered that the native speakers of languages like Japanese show more sensitivity to vowel duration difference changes, and native speakers of English and Mandarin are relatively less sensitive to vowel duration differences. Nevertheless, English speakers have been found more sensitive to vowel duration variants than Mandarin speakers.

In English there is a distinction between lax vowels and tense vowels. These vowels mainly differ in vowel quality, while the role of the vowel duration cue in perception cannot be ignored. One characteristic of tense vowels is that they tend to be longer in duration than lax vowels. For instance, English front high vowels /i/ and /ɪ/ differ in both vowel quality and vowel duration. Duration is an effective cue that adult English speakers make use of when discriminating the two vowels. This kind of duration difference of vowel is defined as “intrinsic vowel duration”, which means that the segment duration is affected by the articulation of the segment itself.

Besides intrinsic vowel duration, there is also “extrinsic vowel duration” in English. In the case of extrinsic vowel duration, the duration of a vowel is affected by articulations of preceding or following segment. One major influence comes from syllable-final consonant voicing. The preceding vowel duration tends to be longer before a voiced consonant than a voiceless consonant. This phenomenon can be observed in many languages, while it is most significant in English. It was reported that English native speakers mainly rely on preceding vowel duration cues in perceiving syllable-final stop voicing (Lawrence 1971; Rodgers 2008).

It is considered that the existence of both intrinsic and extrinsic vowel duration differences in L1 makes English native speakers sensitive to vowel duration variables, and they also use it as a secondary cue in perception. That is the main reason why English native speakers can perceive syllable-final stop voicing even when the final consonant is unreleased.

On the other hand, vowel duration barely plays any role in Mandarin perception.

Although as a tonal language the vowel duration might change due to different tones, previous study has shown that Mandarin native speakers rely less on duration for tone distinction compared to L2 learners (Chang 2011). This could potentially makes it difficult for Mandarin-speaking learners to associate vowel durational information to syllable-final voicing feature in English, which causes the difficulty in perceiving English syllable-final voicing in reduced production variants.

Some of the evidence is provided by Zhang & Hayashi (2015), who compared the perceptual ability of Mandarin native speakers to Japanese native speakers. Japanese is known as an open syllable language. As an open-syllable language, most syllables have a simple CV structure in Japanese. Only in a few circumstances can consonants be permitted to appear at the end of a syllable. For instance, in the case of geminate consonants, the first half /p/, /t/, /k/ and /s/ are permitted to be in final position. However, as mentioned above, since there is phonemic vowel duration in Japanese, vowel duration does play an important role in Japanese perception. This makes them sensitive to the durational differences in spoken language.

Two experiments were conducted by Zhang and Hayashi and the results revealed that Mandarin Chinese speakers performed better than Japanese speakers in general, but their identification rate were more affected by the position of the consonant in the syllable. Japanese speakers can perceive the voicing in unreleased syllable-final consonants better than Mandarin Chinese speakers; also they may make better use of vowel duration as a cue to perceive voicing in syllable-final stops.

How could Mandarin native speakers obtain native-like sensitivity to extrinsic vowel duration differences?

Some studies have shown that the perceptual ability of foreign speech sounds could be improved through training. In the study of Flege & Wang (1989), the performance of English syllable-final /t/ and /d/ by Chinese speakers was improved as the result of feedback training, though the effect is less compared to Cantonese speakers and Shanghai dialect speakers. However, during feedback training, the subjects were exposed to all acoustic features. The improvement in final score is not necessarily related to the increased perceptual ability of making use of preceding vowel duration cues in perception. In the present study, attempts will be made to find out whether or not distributional learning is an effective method for Mandarin native speakers to acquire extrinsic vowel duration.

Some studies show that English native speakers have obtained the sensitivity to extrinsic vowel duration difference at a very early age. Eilers (1977) suggested that infants at around 2 months of age already use vowel duration as a supplementary cue for discriminating final consonantal voicing. Eilers et al. (1984) also reported that 5 to 11 months old infants have the ability to discriminate vowel duration differences, though their performance was much poorer than that of adults. Ko et al. (2009) examined English speakers infants' sensitivity to vowel duration conditioned by syllable-final consonant voicing. The results suggest that infants' sensitivity to extrinsic vowel duration begins to develop between 8 and 14 months.

If acquiring extrinsic vowel duration occurs in the early stage of life, it is possible that the process of acquisition also involves distributional learning. Distributional learning is defined as learning by simply listening to the frequency distributions of the speech sounds in the environment. It is the learning mechanism that human beings use to acquire speech sounds in their early life. It is reported that young infants can acquire phonemic differences through distributional learning (Wanrooij et al. 2014a), and also distributional learning is not limited to linguistic materials but also non-linguistic auditory sequences (Saffran et al. 1999).

Some studies show that distributional learning is also effective for adults. For instance, Ong et al. (2015) reveals that non-tone language listeners can also acquire lexical tone categories through distributional learning. Moreover, it is also demonstrated that distributional learning has long-lasting effects (Escudero & Williams, 2014). Meanwhile, some researches also suggest that the effect of distributional learning in adults is less significant than in infants, indicating that it is a weaker mechanism for learning speech sounds in adults.

Most studies investigated distributional learning at a phonemic level, while the preceding-vowel duration effect in English is phonological. This study will examine whether Mandarin Chinese speakers can obtain sensitivity to extrinsic vowel duration difference through distributional learning. The specific research question is whether training on vowel duration differences is more effective when the syllable-final voicing feature is also available.

Association is the main component of distributional learning, which requires the ability to associate two stimuli that are likely to co-occur. In the case of the present study, the listeners must not only perceive the difference between the long and short vowel, but also build the association between the distributions of vowel duration with the appearance of syllable-final voicing features.

2. Previous Studies

In second language acquisition, there are influences from first language in many different aspects, and the influence can occur in both the perception and production of spoken language.

It has been recognized that the syllable is the basic unit in the perception of spoken language. Previous research suggested that L1 syllable structure has a potential impact on L2 speech acquisition. In Kubozono (1995), two monosyllabic English words CVC and CVC were played to Japanese listeners, and they were asked to make a new word by combining the initial part of the first word and the final part of the second word. The result shows that Japanese speakers prefer to take the CV part from the first word, which means they tend to employ a CV/C segmentation rather than C/VC segmentation pattern.

It is suggested that the L1 segmentation strategy used by learners makes it difficult for them to process within-syllable information in L2, including secondary acoustic cues.

2.1 Syllable-final stop voicing perception by English speakers

In the production of spoken language, the articulation of each sound is not completely isolated; some indirect information could be stored in nearby context, which can serve as a secondary cue in perception when the dominating cue is missing. In the perception of English syllable-final stop consonants, an audible release burst and closure voicing usually signal the voicing feature of the final consonant. At the same time, preceding vowel duration and formant transition can also serve as a secondary cue.

Lawrence (1971) has given evidence that preceding vowel duration is an efficient cue to the perception of syllable-final stop voicing. The study used a series of CVC words ending in either a voiced stop consonant or a voiceless stop consonant, with the preceding vowel duration ranging from 150ms to 350ms. The stimuli were randomly played to 25 English native speakers, and the subjects were asked to make a choice between the members of a minimal pair of words.

The result showed that final consonants are perceived as voiceless sounds when the preceding vowel duration is short and as voiced sounds when preceding vowel duration is long. This indicates that preceding vowel duration is a sufficient cue to the perception of the voicing characteristic of word-final consonants and it is more effective before stop consonants than fricatives. The result also suggests that the perception cued by vowel duration is continuous rather than categorical.

Rodgers (2008) also examined the relative importance of some acoustic cues in determining whether a listener perceives a voiced consonant or a voiceless consonant. The result showed that, among preceding vowel duration, consonants duration and percent glottal pulsing, vowel duration is the strongest of the three features influencing syllable-final stop voicing perception.

Moreover, Hillenbrand et al. (1984) conducted two experiments to investigate among closure voicing, vowel-to-consonant transition, and release burst, what features affect English native speaker's perception.

In the experiments, the subjects were asked to identify the voicing of the syllable-final consonants from the edited CVC words ending in stop consonants: /pɛb/, /pɛd/, /pɛg/, /pag/, /pig/ and /pug/. Vowels for all six syllables are of the same duration. In experiment 1, closure voicing and preceding vowel-to-consonant transition were removed while vowel duration and the final release bursts were retained. In experiments 2, release burst were removed.

It is found that the subjects tended to change from hearing voiced stops to hearing voiceless stops when the closure interval and a portion of the vowel to consonant transition had been removed. And the removal of the final bursts didn't significantly alter the identification. The result indicates that English native speakers do not mainly rely on release burst in perception, and whether final consonant is released or unreleased does not affect English native speaker's perception.

It is not surprising that English native speakers mainly rely on preceding vowel duration in perception. In daily conversation, reduction of sounds is a normal phenomenon and both vowels and consonants can be reduced under some circumstances. It has been stated in some studies that the reduction of consonants is closely related to the position of the sound in a syllabe. Consonant reduction happens relatively easier in syllable-final position than in syllable-initial position (Recasens, D. 2004). In the case of syllable-final stop consonant reduction, cues like release burst may be absent and the reliability is low. Comparing to other cues, preceding vowel duration is more stable and it is less affected by the context. However, this is not the case for Mandarin native speakers.

2.2 Syllable-final stop voicing perception by Mandarin speakers

In Mandarin, there is no contrast between a voiced stop and a voiceless stop, but instead a contrast between a voiceless unaspirated stop and a voiceless aspirated stop. Moreover, stop contrasts occur only in syllable-initial position. Therefore, unlike in English, voicing is not an important feature in the perception of stop consonants in Mandarin. English native speakers usually identify stop categories by voicing categories, while Mandarin speakers are more used to identifying stop categories by aspiration.

Many previous studies have already focused on the perception of English syllable-final consonants by Mandarin Chinese speakers. By comparing their perceptual abilities with English native speakers, we can see that Mandarin speakers use different strategies to perceive syllable-final stop consonants.

Flege (1989) investigated the perception of syllable-final /t/-/d/ contrast by Mandarin Chinese speakers. It is found that Chinese speakers can perceive an unedited English /t/ and /d/ in syllable-final position very well, but when the release burst and closure voicing were removed, they performed much worse than English native speakers. The result indicated that Chinese speakers are not familiar with syllable-final stops and they attempted to rely on syllable-initial cues to identify syllable-final /t/ and /d/.

Flege & Wang (1989) conducted another experiment to examine Chinese speakers' perception of the syllable-final /t/- /d/ contrast in English, and this time the influence of dialects was also taken into consideration. The study attempted to find out if the syllable structure of one's native language is related to the learner's sensitivity to non-native sounds appearing in certain positions. Although a contrast between /t/ and /d/ only exists in the initial but not the final position of Mandarin words, Cantonese permits unreleased /p/, /t/, /k/ in word-final position, and in the Shanghai dialect word-final glottal stop is permitted. So the experiment was conducted by dividing all the subjects into three groups based on their language background: Mandarin speakers, Cantonese speakers and Shanghai dialect speakers. Since the number of consonants that are permitted to appear in syllable-final position in their first language is different, the three groups of subjects were presumed to show a different sensitivity to the perception of stop consonants in a second language.

The result showed that the correct identification rate of native speakers of Cantonese, as predicted, was the highest at 88%, and Shanghai dialect speakers had a higher correct identification rate than native Mandarin speakers (79% and 69% respectively).

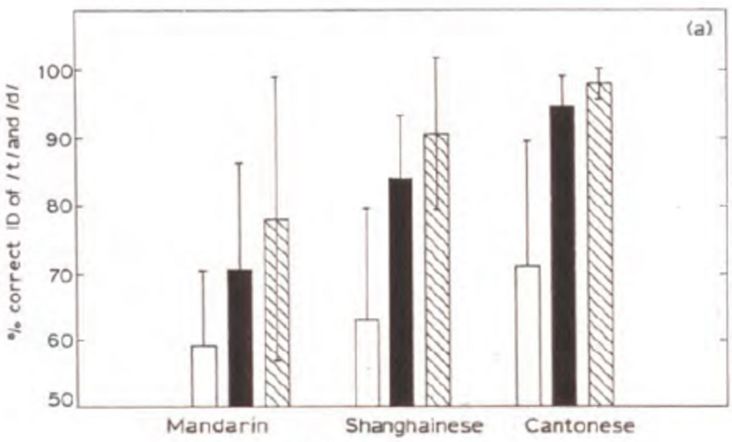


FIGURE 1 | The mean percentage of correct identification of /t/ and /d/ by Mandarin, Shanghai dialect and Cantonese speakers before □, during ■, and after ▨ feedback training. (Flege & Wang, 1989)

The study also investigated the effect of feedback training on perception. During the training phase, the subjects were informed of the correct answer soon after they made the judgment. As the result of training, the performance of each group was improved. But the difference between pre-training phase and post-training phase was greater for Cantonese and Shanghai dialect speakers (26% and 28%) than for Mandarin speakers (19%).

2.3 Phonemic vowel duration for Mandarin speakers

It was proved by previous studies that vowel duration is the most effective cue to the perception of stop consonant's voicing feature for English native speakers. Whether Mandarin speakers can make good use of this cue becomes the next question to investigate. The key point is considered to be their sensitivity to the change of vowel duration and by reviewing some previous studies, we can see there is a big difference between Mandarin speakers and English native speakers in this aspect.

As mentioned above, vowel duration difference is not phonemic in Mandarin. Some might argue that this is not entirely true, because some tonal differences also reflect in durational differences. It is known that there are in total five tones (including neutral tone) in Mandarin and tone 2 and tone 3 have been considered to be the most perceptually confusable for non-native speakers. Simply put, Mandarin tone 2 is characterized as having a high-rising pitch contour and tone 3 is characterized as having a low falling-rising pitch contour.

Chang (2011) investigated whether vowel duration serves as a perceptual cue and whether L1 and L2 listeners weigh the cue in the same manner in distinguishing the two tones. The experiment consisted of two forced choice discrimination tasks. Task 1 used the stimuli where tone 3 was naturally longer than tone 2, while Task 2 used duration-normalized stimuli. The result showed that L2 listeners relied more on duration in the discrimination of tone 2 and tone 3, while Mandarin native speakers gave less perceptual weight to duration.

The result supports the fact that vowel duration barely plays a role in Mandarin perception. At the same time, efforts have been made to examine how Mandarin speakers acquire a foreign language showing phonemic vowel duration distinction.

Kurihara et al. (2006) conducted a listening test to investigate how Mandarin native speakers perceive Japanese long vowels by vowel duration. The result showed that Mandarin speakers perceive some vowels with short duration as long vowels and it was difficult to see a similar tendency of categorical perception like Japanese native speakers from them. The result was then compared to their Japanese ability levels and no relations were found.

The lack of vowel duration difference in Mandarin also reflects in prosody production. Minagawa-Kawai (2001) compared the production of nonsense words by Mandarin native speakers and Japanese native speakers. As is well known, Japanese is a mora-timed language, which means the duration of every mora is equal. That is to say, the duration of the consonant and the vowel within a mora usually have a compensating relationship. It was hypothesized in the study that Japanese speakers would show a stronger tendency to keep CV length equal than speakers of other languages, because CV must not be too long in Japanese where CV and CVC/CVV have a durational contrast.

Previous studies already examined the existence of temporal compensation while Minagawa-Kawai focused on the issue of how differently temporal compensation is performed depending on language. The study tested the hypothesis by comparing Japanese with Mandarin, which has a different rhythm pattern from Japanese. In the pronunciation test, two nonsense words "maCVka" and "maCVnka" were made, where the CV is /ra/, /ba/, /sa/, /ri/, /bi/ and /si/, and these words are inserted in a carrier sentence.

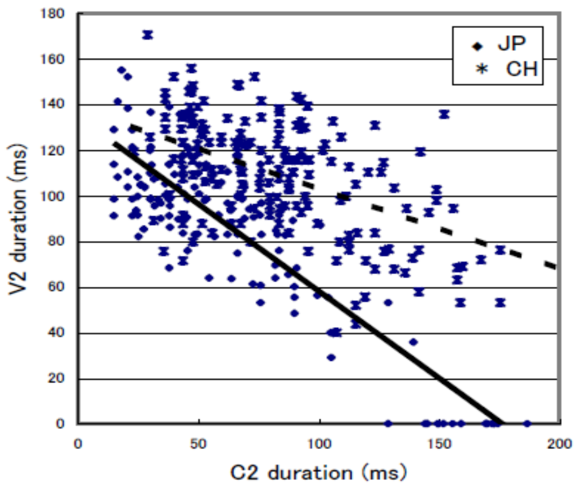


FIGURE 2 | Correlation between duration of C_2 and V_2 for Chinese and Japanese. (Minagawa-Kawai 2001)

The records were analyzed on computer and segmental durations of the recorded utterances were measured. The duration of C_2 (e.g. /b/ in /mabaka/) and the duration of its following vowel V_2 in each utterance were plotted.

Fig 2 compared the tendency of temporal compensation effect across language. From the chart we can see that when the consonant duration gets longer, the vowel duration gets shorter. This is true for both Japanese and Chinese speakers but it is clear from the figure that the extent of temporal compensation is larger for Japanese than for Chinese. Japanese speakers show stronger temporal compensation because CV and CVV have a durational contrast, while the contrast doesn't exist in Mandarin.

Although this test examines temporal compensation in production, the result also gave some evidence that Mandarin speakers pay less attention to vowel duration in spoken language than Japanese speakers.

2.4 Vowel duration as secondary acoustic cue for Mandarin speakers

Zhang & Hayashi (2015) conducted two experiments to examine the perception of English consonants in syllable-final position by Mandarin native speakers and Japanese native speakers, and to also find out to what extent they rely on the vowel duration cue to perceive syllable-final stop voicing.

The stimuli used in this experiment were nonsense English words containing 13 American English consonants. All words have the same structure of $C_1V_1C_2C_3V_2$, with /e/ and /i/ in V_1 and V_2 position and /p/ in C_1 position. The 13 target consonants appear in C_2 and C_3 position alternately. Among them, words with same consonant in C_2 and C_3 positions were removed from the list. An American English native speaker recorded the stimuli in a soundproof room. The speaker pronounced all the words naturally and the consonants in C_2 position were all unreleased.

The subjects were told to listen to a series of nonsense English words with a structure of /peC_2C_3i/ and they were asked to concentrate on the two consonants in C_2 and C_3 positions, and then to write down the sounds they had identified on the answer sheet. The results showed that although Chinese subjects have a higher average score than Japanese subjects in general, the perception of Chinese subjects is more affected by syllable position.

It seemed that when a consonant appears in the onset position, Chinese speakers do not have any problems identifying it, but when it appears in the coda position, the identification rate could be extremely low. For example, Chinese speakers have an average identification rate of 95.83% for word-initial /t/ while for word-final /t/ this number is only 15%.

And also, as shown in Fig. 3, when it comes to syllable-final stop consonants and fricative consonants, Chinese speakers made more mistakes than Japanese speakers in perceiving the voicing feature. This difference suggests that Chinese subjects must be less sensitive to certain acoustic features cuing voicing than the Japanese subjects, which was further investigated in experiment 2.

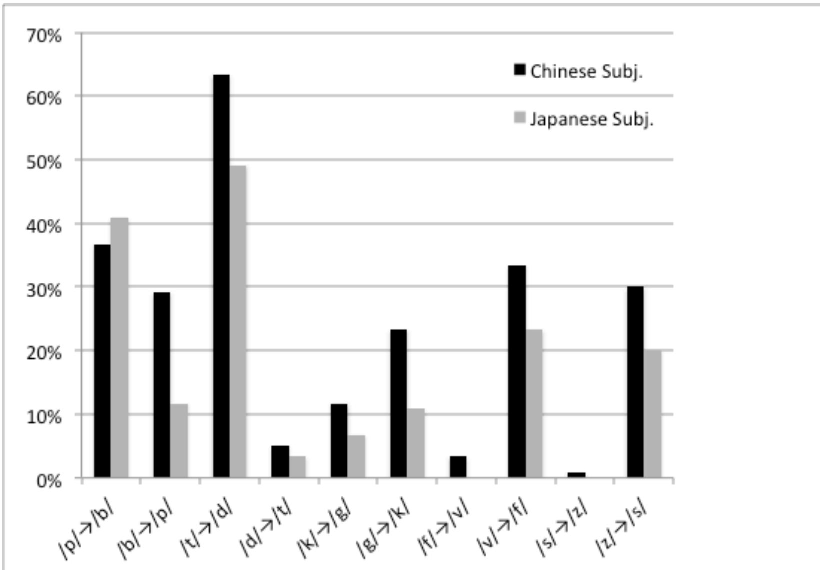


FIGURE 3 | Percentage of errors occurring within voiced-voiceless contrast in syllable-final position. (Zhang & Hayashi, 2015)

A second experiment was also conducted in the study. Six English words with a simple CVC structure /cɑp/, /cɑb/, /bɛt/, /bɛd/, /pɪk/ and /pɪg/ were recorded by an American English native speaker. For each word, the vowel durations were measured. The stimuli were made by editing the words ending in voiceless stops. The release burst was eliminated and the duration of the preceding vowel was lengthened in five steps. In the final step, the duration is equal to that of the preceding vowel of the words ending with voiced consonants. The subjects were asked to identify whether the word ended with a voiced or a voiceless consonant.

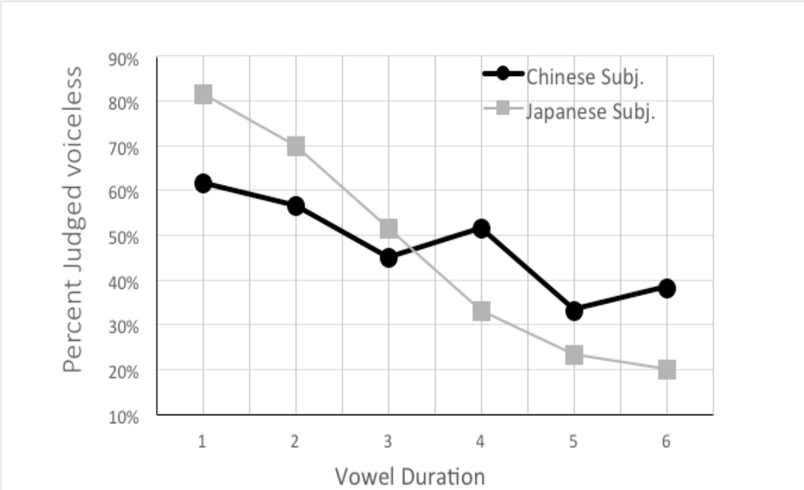


FIGURE 4 | Percentage judged as voiceless by Chinese speakers and Japanese speakers with the change of vowel duration. (Zhang & Hayashi, 2015)

The results showed a significant difference between Mandarin and Japanese speakers (Fig. 4). In Figure 4, the percentage judged as voiceless by Japanese speakers gets lower if the preceding vowel duration gets longer. There was a significant difference between groups of different vowel durations (ANOVA, $p < 0.001$). This suggested that for the Japanese subjects, the final consonant was likely to be perceived as voiceless when the preceding vowel duration is shorter. This is the same result as reported for English native speakers by Lawrence (1971). Thus Japanese speakers show a tendency to use vowel duration as a voicing cue in way similar to English speakers. No significant difference was found as a function of vowel duration in the perception of Mandarin native subjects (ANOVA test, $p=0.26$).

It was found in the interviews after the experiment that most Japanese subjects can notice the differences between each stimulus and they have a clear standard to help them to make a choice. Some said they have a clear standard for making the judgments, they chose the voiceless sound when the vowel sounded shorter and chose the voiced sound when the vowel sounded longer. This identification strategy is thought to come from their native language. In Japanese there is a contrast between long vowels and short vowels, and also a contrast between words with geminates and words without geminates. It is known that geminates can change the duration of the preceding vowel and even the posterior vowel. This is thought to be the reason why Japanese speakers can notice the change of vowel duration easily. On the other hand, most Chinese speakers said that they cannot tell the difference between the stimuli in each group when they were asked how they made their choice. Some said that they could perceive the difference to some extent but they could not tell exactly what part is different.

This finding may account for their relatively poor performance in Experiment 1 for identifying voicing features of syllable-final stops. This is considered to be a negative transfer from their first language. As was discussed above, vowel duration difference is phonemic in Japanese but not in Mandarin Chinese.

From above, it is clear that both in the perception of Mandarin and foreign languages, Mandarin native speakers do not regard vowel duration as an effective cue. Besides the difficulties in perception, Mandarin speakers show difficulties in production as well. It is reported that Mandarin native speakers often omit an English final consonants or insert a vowel after the final consonant (Flege, 1989). Xu & Demuth (2012) provided an acoustic analysis of English syllable-final consonant production by Mandarin leaners of English. A total of eight female L1 speakers of Mandarin participated in the study. The target consonants were embedded in four monosyllabic CVC words. As a result, no significant difference was found in preceding vowel duration between voiced and voiceless stops.

Flege et al. (1992) examined the production of word-final English /t/ -/d/ contrast by native speakers of English, Mandarin and Spanish. In the study, English native speakers were asked to identify the voicing feature in syllable-final stops produced by talkers in five groups: English native speakers, experienced and inexperienced Spanish-speaking English learners, and experienced and inexperienced native Mandarin speaking English learners.

As a result, the experienced second language learners' stops were correctly identified less often than stops produced by the native English speaker, and the difference is significant. No significant difference in identification rate was found between the stops produced by experienced second language learners and the stops produced by the inexperienced L2 learners. Moreover, it is also revealed that the productions of native English speakers show some significant acoustic features. For instance, they produced vowels significantly longer before /d/ than /t/. Although the same kinds of acoustic differences between /t/ and /d/ were found also found in the production of the L2 learners, the difference is relatively smaller. Finally, there is a strong correlation between vowel duration differences and the listeners' identification under edited pronunciation condition (where post-vocalic cues were removed) than under full pronunciation condition.

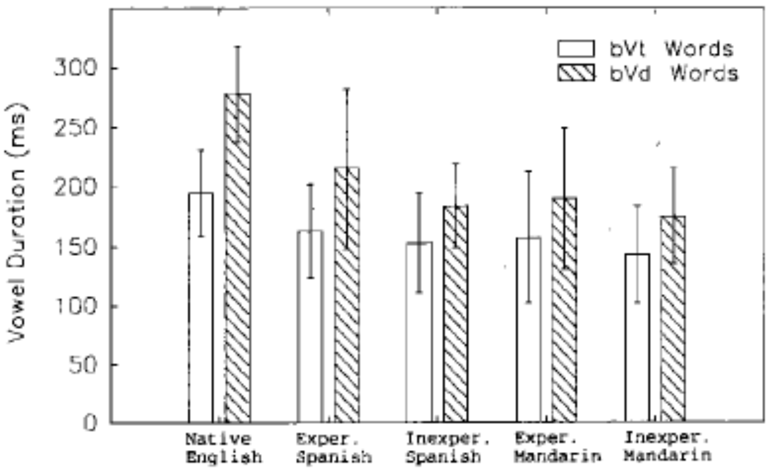


FIGURE 5 | Mean duration of vowels (ms) in minimally-paired words ending in /d/ and/ t/ spoken by subjects in five groups. (Flege et al., 1992)

This study not only supports that English native speakers rely on vowel duration to perceive syllable-final voicing, but also reveals that the preceding-vowel duration effect in Mandarin speaker’s production is smaller compared to English native speakers.

In second language acquisition, production and perception are closely connected. The fact that Chinese speakers could not make use of the vowel duration cue efficiently and also they could not produce vowel duration cue in their own pronunciation added another piece of evidence for the statement. This will potentially cause misunderstanding in their communication with English native speakers.

2.5 Previous studies on distribution learning

When and how do native speakers acquire the sensitivity to extrinsic vowel duration difference?

Ko et al. (2009) examined 8-month-old and 14-month-old infants' perceptual sensitivity to vowel duration conditioned by post-vocalic consonantal voicing. In the experiment, three CVC minimal pairs /bæg/-/bæk/, /cʌb/-/cʌp/ and /pɪg/-/pɪk/ were used as the stimuli; half the infants heard CVC stimuli with short vowels, and half heard stimuli with long vowels. In both groups, stimuli with voiced and voiceless final consonants were compared. The result showed that older infants are sensitive to mismatching vowel duration and consonant voicing in the short condition, while no significant result was found in younger infants in either condition. The study suggested that sensitivity to preceding vowel duration conditioned by post-vocalic consonantal voicing might develops over the course of the second half of the first year of life. Although it is still not clear how the sensitivity is obtained, it is possible that distributional learning also involved in the process of acquisition.

Distributional learning is defined as learning by simply listening to the frequency distributions of the speech sounds in the environment. It is the learning mechanism that human beings use to acquire speech sounds in their early life.

Wanrooij et al. (2014a) investigated the effect of fast distributional learning in the lab. In the study, Dutch infants aged 2 to 3 months were presented with either a unimodal or a bimodal vowel distribution based on the English /ɛ/-/æ/ contrast for only 12 minutes. After the training, their ability to discriminating discriminate [æ] from [ɛ] were tested in an ERP oddball paradigm. The stimuli were synthesized in Praat and they only varied in the values of F1 and F2. F3, F4 and F5 were set to the same value for each stimulus. The duration of each stimulus was also kept at 100ms.

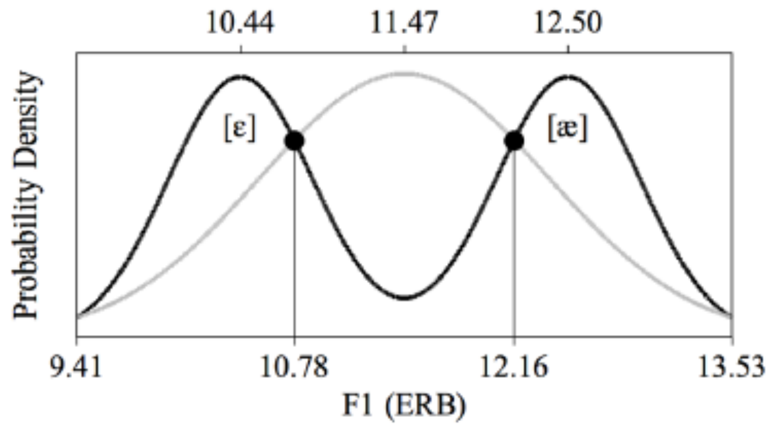


FIGURE 6 | Unimodal (gray curve) and bimodal (black curve) training distributions of the first vowel formant (F1). The values of the test stimuli lie at the intersections of the two distributions. (Wanrooij et al., 2014a)

If distributional training is effective, infants trained bimodally should have a better performance in discriminating the two test stimuli. The result shows that the infants exposed to bimodal distribution can better discriminate /ɛ/-/æ/ contrast after the training than the infants exposed to unimodal distribution.

Wanrooij et al. (2014b) further examined whether the capacity for using the mechanism is different in adults than in infants. In the experiment, the participants were Dutch native speakers that had been raised monolingually. The subjects were exposed to either a bimodal distribution that suggested the existence of the two English vowels, or to a unimodal distribution that did not. Before and after exposure the participants were tested on their discrimination of a representative [æ] and a representative [e], by measuring mismatch responses (MMRs). As a result, the significant effect of bimodal distribution in infants observed by Wanrooij et al. (2014a) was not shown in the adult subjects.

Ong et al. (2015) examined whether non-tone language listeners can acquire lexical tone categories through distributional learning. In the study, Native Australian English listeners were trained on a Thai lexical tone minimal pair. During Training, the subjects either heard a unimodal distribution that would induce a single central category, or a bimodal distribution that would induce two separate categories. A discrimination task was conducted before and after training. It is found that the bimodal group outperformed the unimodal group at Post-test relative to at Pre-test.

Furthermore, Escudero & Williams (2014) examined both short- and long-term effects of distributional learning of phonetic categories on non-native sound discrimination over a 12-month period. Two groups of listeners were exposed to a two-minute distribution of auditory stimuli in which the most frequently presented tokens either approximated or exaggerated the natural production of the speech sounds, whereas a control group listened to a piece of classical music for the same length of time. Discrimination by listeners in the two distribution groups improved immediately after the short exposure, replicating previous results. Crucially, this improvement was maintained after six and 12 months, demonstrating that distributional learning has long-lasting effects.

Pająk & Levy (2012) conducted an experiment on adults with different language backgrounds. In the study, native speakers of Korean and Mandarin were tested. In Korean, duration is a cue to distinguishing phonetic categories (e.g., [pul] 'fire' vs. [pu:l] 'blow'). Therefore they are better at discriminating duration contrasts than are speakers of Mandarin, in which there are no length contrasts. While Mandarin speakers are considered to be better in distinguishing place of articulation contrast between alveolo-palatal and retroflex sibilants. Mandarin speakers and Korean speakers participating in the experiment they were exposed to the sound distribution in a hypothetical language that could be interpreted as either a place distinction (alveolopalatal vs. retroflex) or a duration distinction (short vs. long).

The results showed that the effect of training on length is better for Korean speakers, which suggests that speakers of Korean were biased toward inferring a length-based category distinction and against inferring a place-based category distinction. On the other hand, the type of training did not have a clear effect on Mandarin speakers.

In this study, an experiment was conducted to examine whether Mandarin Chinese speakers can obtain sensitivity to extrinsic vowel duration difference through distributional learning. The specific research question is whether distributional training on vowel duration differences is more effective when syllable-final voicing feature is also available during the training.

3 Rationale

The present study aims at finding an effective training method to increase Mandarin speaker's perceptual ability towards syllable-final stop voicing.

In the study of Flege & Wang (1989), Mandarin speakers' perception of syllable-final /t/-/d/ contrast was investigated. The study also examined the effect of feedback training. During the feedback-training phase, the subjects were informed of the correct answer immediately after they made the choice each time. The speakers' performance in the pre-training phase was then compared to the during-training phase and the post-training phase. As a result, the performance of English coda /t/ and /d/ by Chinese speakers was improved as the result of feedback training, although the improvement is small compared to Cantonese and Shanghai dialect speakers. However, during feedback training, the subjects were exposed to more than one acoustic feature. The improvement in perception is not necessarily related to the increased ability of making using of preceding vowel duration cue in perception.

Compared to feedback training, distributional learning can better make the subjects focus on one specific acoustic feature; this might potentially increase the effect of training. Therefore, it is necessary to examine whether the sensibility of Chinese speakers towards extrinsic vowel duration can be obtained through distributional training. The specific research question is whether the presence of syllable-final voicing feature can enhance the effect of training of vowel duration differences.

4 Experiment

A perceptual experiment was conducted to examine whether or not Mandarin speakers can acquire the effect of preceding-vowel duration through distributional learning. In order to investigate the effect of distribution learning, the experiment consists of three phases: pre-test phase, training phase and post-test phase.

4.1 Subjects

40 Mandarin Chinese native speakers participated in the experiment. All subjects reported normal hearing and all had learned English as an L2 after the age of 14. Cantonese dialect speakers and Shanghai dialect speakers were not included.

All subjects receive the same discrimination test in the pre-test phase and post-test phase. In the training phase, they were divided into four groups: released bimodal group, released unimodal group, unreleased bimodal group and unreleased unimodal group. The type and distribution of the stimuli that a subject was exposed to were different for each group.

4.2 Stimuli

There were two types of stimuli in the experiment, for the released bimodal and unimodal groups the stimuli were from a continuum changing from [mat̚_^{t}] to [ma:d̚_^{d}]. For the unreleased bimodal and unimodal groups, the continuum changing from [mat̚] to [ma:d̚] was used. The only difference between the two sets of stimuli is whether the final stop consonant is released or not.

The sources used to create the stimuli were recorded in a soundproof laboratory, which are two CVC words /mat/ and /mad/. The recordings were then edited with Praat (version 6.0.21). Two types of stimuli are needed for the experiment: the released stimuli and the unreleased stimuli. Recordings with a steady pitch contour were chosen as the released stimuli. The unreleased stimuli were made by eliminating the syllable-final stop consonants /t/ and /d/.

Vowel duration was the parameter in this experiment, and it should be shorter when preceding a voiceless stop consonant and longer when preceding a voiced stop consonant. The duration was controlled by shortening the duration of the vowel in /mad/ by several steps. In order to keep a smooth pitch transition after the editing, every step one period from each steady part of the vowel was cut off. At the same time, cutting two adjacent periods was avoided during the editing. In total, there were 18 released stimuli and 17 unreleased stimuli. Table 1 shows the vowel duration of the stimuli used in the experiment. In the experiment, the Stimuli were presented to the subjects in Praat.

Stimulus	Vowel Duration (ms)
a-4	345
a-5	335
a-6	325
a-7	315
a-8	305
a-9	295
a-10	285
a-11	275
a-12	265
a-13	255
a-14	245
a-15	235
a-16	225
a-17	215
a-18	205
a-19	195
a-20	185

Table 1 | Vowel duration of the stimuli used in the experiment.

4.3 Procedures

Pre-test phase and Post-test phase

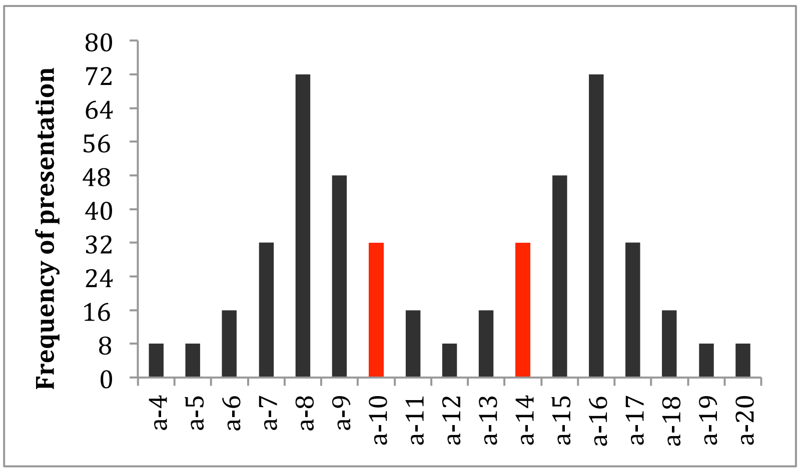
During the pretest and posttest phase, all subjects received the same discrimination test. In the test, the subjects were presented with two stimuli /ma/ with different vowel durations (245ms and 285ms) each time. They were then asked to listen to two words from a new language, and to determine whether the two words they heard were the same. After the subjects made the judgment, the next pair of stimuli was presented automatically. This procedure was repeated 120 times, each time in a randomly generated order (combinations of two same words are also included).

The two test stimuli were chosen because they had been presented in the same frequency in both bimodal distribution and unimodal distribution during the training phase. Therefore, the four groups of subjects came to the test phase with equal prior exposure in the training phase to sounds in the region of the test stimuli, so that any difference between the groups observed in the test could not be attributed to differences in familiarity with the test stimuli.

Training phase

In the training phase, the subjects were divided into four groups: released bimodal group, released unimodal group, unreleased bimodal group and unreleased unimodal group.

Stimuli used for the unreleased bimodal group and the unreleased unimodal group were from the [mat̚] - [ma:d̚] continuum with vowel duration ranging from 185ms to 345ms. Stimuli used in the released bimodal group and released unimodal group were from the [mat̚_^{t}] - [ma:d̚_^{d}] continuum with vowel duration ranging from 185ms to 345ms and /C/ changing from fully released /t/ to fully released /d/ when the vowel duration increased to 265ms (stimulus a-12). The stimuli were created in this way so that voiceless final /t/ always occurs with relatively shorter vowels and voiced final /d/ always occurs with relatively longer vowels.



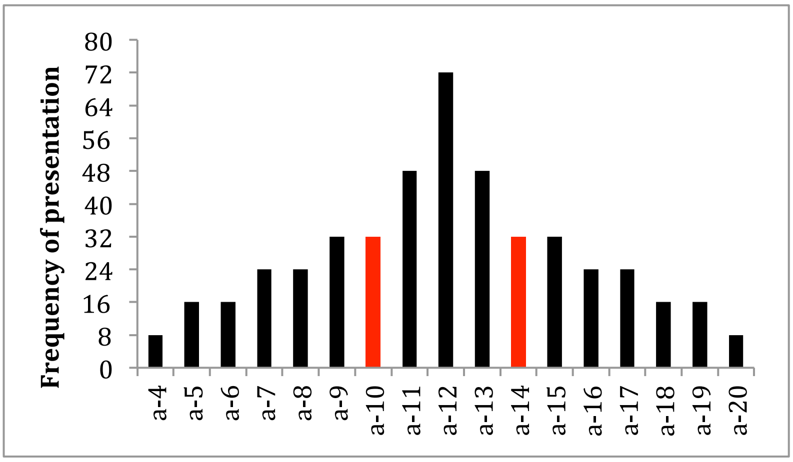


Figure 7 | Frequency of presentation of stimuli in bimodal groups (up) and unimodal groups (down). The values of the test stimuli are indicated in red color.

During the training, each stimulus was repeated multiple times. The total number of presentations of stimuli was 472 for all groups, while the frequency distribution of the stimuli was different in the bimodal group and the unimodal group. The unimodal groups induced a single central category, while the bimodal groups induced two separated categories. Figure 7 shows the frequency of presentations of stimuli in bimodal groups (top) and unimodal groups (bottom). In bimodal distribution, the stimuli a-8 and a-16 occur the most frequently, which is 72 times each. The stimuli a-4, a-5, a-12, a-19 and a-20 occur the least frequently, which is 8 times each. By contrast, the unimodal mean lay exactly in the middle of the range of vowel duration values and in between the two bimodal means. The most frequently occurring stimulus was the a-12 and the least frequently occurring ones were a-4 and a-20. The stimuli were randomized and presented with Praat based on two different distribution scripts. The order of presentation of the stimuli in the training was randomized separately for each subject. The subjects were told to listen to a series of words, and during the training they were not required to make any operations on the computer.

The scores of each subject in pre-tests and post-tests were calculated. Only comparing the result of released bimodal group and released unimodal group could not lead to a conclusion on whether the training is effective for extrinsic vowel duration difference acquisition. The score difference does not necessarily suggest that the subjects can better associate the vowel duration differences with syllable-final voicing. It might be simply because that the training phase increased the subjects' sensitivity to vowel duration. In order to exclude the possibility, the result of the groups trained with released stimuli will be compared to that of the groups trained with the unreleased stimuli. In those two groups, the training only aims at increasing sensitivity to phonemic vowel duration difference. This is why the two unreleased groups are necessary in this experiment

The specific research question of this study is whether distributional learning on vowel duration differences is more effective when syllable-final voicing feature is also available in the stimuli. In the experiment, the bimodality effect in both released and unreleased condition was examined. The main point we want to see is whether or not the bimodality effect is bigger under the training with released syllable-final consonants than under the training with unreleased syllable-final consonants. Therefore, besides the difference between pre-test and post-test and the difference between bimodal and unimodal, the difference between released and unreleased is also an indispensable factor in this study. We expected to see a statistically significant triple interaction among test condition, distribution type and stimulus type. Assuming that the distribution type (bimodal vs. unimodal) has an effect on the score across different test condition (pre vs. post), this interaction could reveal that stimulus type (released vs. unreleased) has an effect on this "distribution type effect". Only when this interaction is significant, we can say that training for duration is more effective with released stimuli.

4.4 Result

Table 2 gives the pre- and post-test percentages correct (i.e., the percentage of correct judgment of the 120 test stimuli) and the difference (i.e., the post-test minus pre-test percentage correct) for all four groups.

An ANOVA on all pre-test scores did not display a significant difference between the four groups ($p = 0.74$). This suggests the equality of the groups before training.

The difference between pre-test and post-test score is a measure of improvement after training. For the subjects exposed to released stimuli of a bimodal distribution, score in the post-test was improved compared to that in the pre-test, and the score difference is significantly different from zero (t-test, $p<0.01$). There is no significant improvement in the subjects from the released unimodal group (t-test, $p = 0.85$). An ANOVA with difference scores as the dependent variable revealed a significant difference between the released bimodal group and the released unimodal group ($p = 0.02$).

	pre	post	Difference
Released Bimodal	54.24 (4.80)	62.25 (8.36)	8.00
Released Unimodal	58.42 (10.20)	59.32 (12.08)	0.91
Unreleased Bimodal	58.17 (10.84)	62.5 (12.34)	4.33
Unreleased Unimodal	56.47 (10.72)	56.67 (10.58)	0.25

Table 2 | Pre- and post-test percentages correct, and difference. Standard deviations between participants in each group are given between parentheses.

On the other hand, the score difference for unreleased bimodal group is not significantly different from zero, and also ANOVA test with difference scores as the dependent variable revealed no significant difference between the unreleased bimodal group and the unreleased unimodal group ($p = 0.09$). That is to say, compared to training with unreleased stimuli of a unimodal distribution, training with bimodally distributed stimuli did not significantly increase the subjects' sensitivity towards vowel duration difference. Moreover, there is a main effect of test type, and the odds of post-test people score was improved by 14 percent compared to the odds of pre-test people ($p = 0.0016$). There is also an interaction effect of test type and distribution type. The average of pre-test unimodal people and post-test bimodal people scores 24.9 percent better than the average of pre-test bimodal people and post-test unimodal people ($p = 0.0055$).

However, no significant triple interaction was found among distribution type, test type and stimulus type ($p = 0.83$). For that reason, the possibility that the score improvement observed in released bimodal group is only in the phonemic level cannot be excluded. Therefore, we cannot draw a conclusion that training with released stimuli is more effective.

4.5 Discussion

In order to investigate the effectiveness of distributional learning of extrinsic vowel duration difference, an experiment was conducted on Mandarin Chinese native speakers. In the experiment, the subjects were divided into four different groups: released bimodal group, released unimodal group, unreleased bimodal group and unreleased unimodal group. The result shows significant improvement before and after training in released bimodal group. However, training with released stimuli did not significantly improve the subjects' perception than training with unreleased stimuli.

As mentioned above, bimodal released group is the main group investigated. The result showed that the subjects trained in this group indeed had a better performance in post-tests. However, the improvement observed in released bimodal group could not lead to the conclusion that the training is effective for acquiring extrinsic vowel duration differences. The possibility that the training only improved the sensitivity to phonemic vowel duration difference must be excluded.

For the subjects in unreleased groups, the result revealed no significant difference between unreleased bimodal group and unreleased unimodal group. The subjects trained with bimodally distributed stimuli were not significantly better in identify vowel duration differences after training than the subjects trained with unimodal distributed stimuli.

It turned out that simple training in a phonemic level did not help Mandarin speakers to become more sensitive to vowel duration differences. Although during the training with unreleased stimuli, the subjects only need to focus on the durational information. It seems that this is not easier for Mandarin native speakers. It is possible that under training with released stimuli, it is easier for the subject to associate the durational change with the final consonant, which enhances the learning effect.

Most subjects reported the experiment as a difficult task. In the pre-test phase, some Mandarin Chinese speakers reported that duration difference was not taken into consideration in discriminating the two words; some reported that they could not accurately perceive the durational difference. Although after the training they were able to notice that the difference lies in duration, making correct judgment is still a hard task for them. This might partially be because of the choice for test stimuli. In the discrimination test, stimuli with vowel duration 245ms and 285ms were chosen as the test stimuli. It is possible that the durational difference between the two stimuli is too small while the effect of training is limited.

Language transfer is one of the earliest theories of second language acquisition. In the study of phonological acquisition, the effect of this transfer is divided into two categories. Positive transfer occurs when the L1 and L2 patterns are similar, and habits from the L1 are carried over to the L2. Negative transfer occurs when the L1 and L2 patterns are different, and habits carried over interfere with the L2. In the acquisition of vowel duration length effect by Mandarin speakers, negative transfer from L1 occurs. The speech learning model theory (Flege, 1995) has indicated that the phonetic system of native and foreign sounds can affect each other. The hypothesis stated that so long as the learner can perceive part of the different features of a foreign sound, there would be more chances for a new category to be formed, which does not belong to any native sound category. The more different a foreign language sound is from native sounds, the easier it is for learners to identify it. And if two sounds are not in a phonemic contrast with each other in one's native language, it will be very difficult for the listener to tell them apart, even in a foreign language. This hypothesis is also supported by the report by Aoyama et al. (2004) who stated that it is easier for Japanese speakers to improve through perceptual training on /r/ sounds than /l/ sounds, since /r/ sounds do not exist in Japanese and they have more specific features.

This provides a possible explanation to the invalid training of phonemic vowel duration on Mandarin speakers. Since long vowel and short vowel is not a phonemic contrast in L1 for Mandarin speakers, training on durational differences brings more difficulties than training on complete novel sounds.

Wanrooij et al. (2014b) has indicated that distributional learning is less effective in adults than infants. It remains to be seen whether the same type of training can help infants at an early age to acquire extrinsic vowel duration differences. The experiment design also has potential influence on the training effect, because a pre-test would have been an additional distributional training and this could have distorted the intended training distributions. This also needs to be taken into consideration in further experiment design. It can be seen from the result that there is a big personal difference between the participants; more participants are needed in order to achieve a more accurate result.

In the perception of syllable-final stop voicing, formant transition can also serve as an acoustic cue. In recent years, the transition pattern of the first formant (F1) in preceding vowel also starts to draw people's attention. The first formant (F1) of vowels has been seen to vary with the voicing of the following consonant, i.e. F1 offset frequency of the preceding vowel usually has a lower value when it is followed by a voiced stop consonant. Previous studies have shown that English native speakers can also rely on vowel to consonant transition to perceive syllable-final voicing (Hillenbrand et al., 1984). Further studies will be conducted to investigate whether distributional learning can help Mandarin speakers to acquire the ability to use formant transition cues in syllable-final voicing perception. And also, future research is necessary to examine the persistence of short-term distributional learning over time.

In second language acquisition, production and perception are closely connected. And the fact that Chinese speakers could not make use of the vowel duration cue efficiently and also they could not produce the vowel duration cue in their own pronunciation added another piece of evidence for the statement. This will potentially cause misunderstanding in their communication with English native speakers. Previous studies indicated the possibility of improving learners' pronunciation of foreign sounds through training. The next step is to examine whether or not the improvement in production can in turn improve the perceptual ability.

5 Conclusion

This study investigated the perception of syllable-final stop voicing by Mandarin native speakers. This is one of most common difficulties that Mandarin native speakers encounter during English learning. It is considered to be due to negative transfer from their first language. Besides the reason that voicing contrast in syllable-final position is unfamiliar to Mandarin speakers, the fact that Mandarin lacks intrinsic and extrinsic vowel duration differences also makes it difficult for Mandarin native speakers to perceive durational difference information in a foreign language. Therefore Mandarin speakers have more difficulties in perceiving syllable-final consonant voicing when the final consonant is unreleased.

The specific research question was whether distributional learning on vowel duration differences is more effective when syllable-final voicing feature is also available. In order to achieve the research goal, an experiment was conducted on 40 Mandarin native speakers.

The result revealed a significant difference between released bimodal group and released unimodal group. At the same time, no significant triple interaction was found among distribution type, test type and stimulus type. In other words, although there was a significant improvement in the released bimodal group, we cannot draw a conclusion that training is more effective with released stimuli. Further research needs to be done to confirm that the improvement in the released bimodal group is due to the factor of syllable-final voicing.

Foreign language skill has always been seen as an important part in education in China. Every year new foreign language schools are being established. But due to the lack of phonetic studies, the pronunciation training is still insufficient. As one of the problems that haven't been widely noticed, the voiced and voiceless contrast has confused many English learns in China. The present study will hopefully help them to understand this issue fundamentally and draw more people's attentions to the importance of phonetics in English education.

Acknowledgments

First and foremost, I would like to express my gratitude to my supervisor Paul Boersma for the precious comments, remarks and engagement through the whole process of this master thesis. I also want to thank the participants who have willingly taken their precious time to take part in the experiments. Many thanks to my friends who have discussed the topic with me since the very beginning and have given many useful advice and technical supports. Finally, thanks to my parents for their continuous and unconditional supports. This thesis would not have been possible without the support and helps of the people that mentioned above.

REFERENCES

[1]. Aoyama, K. et al. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. Journal of Phonetics, 32, 233–250.

[2]. Boersma, P., Weenink, D. (2006). Praat: Doing Phonetics by Computer. (Version 5.4.04). [computer software]. Retrieved from http://www.praat.org/.

[3]. Chang, Y.S. (2011). Distinction Between Mandarin Tones 2 and 3 for L1 and L2 Listeners. Proceedings of the 23rd North American Conference on Chinese Lingusitics (NACCL-23), 2011. Volume 1, edited by Zhuo Jing-Schmidt, University of Oregon, Eugene. Pages 84-96.

[4]. Eilers, R. (1977). "Context-sensitive perception of naturally produced stop and fricative consonants by infants," J. Acoust. Soc. Am. 61, 1321–1336.

[5]. Eilers, R., et al. (1984). "The discrimination of vowel duration by infants," J. Acoust. Soc. Am. 75, 1213–1218.

[6]. Escudero, P., Williams, D. (2014). Distributional learning has immediate and long-lasting effects. Cognition. 2014 Nov; 133(2): 408-13.

[7]. Flege, J.E. (1995). Second Language speech learning: Theory, findings, and problems. In W. Strange (ed.), Speech perception and linguistic experience. pp. 233-273. Baltimore: York Press.

[8]. Flege, J.E. (1989). The perception of word-final /t/ and /d/ by native and Chinese listeners, Journal of the Acoustical Society of America, to appear.

[9]. 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.

[10]. 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. Journal of the Acoustical Society of America 92, 128-142.

[11]. Hillenbrand, J., Ingrisano, D.R., Smith, B.L., Flege, J. E. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. The Journal of the

Acoustical Society of America 08/1984; 76 (1): 18-26.

[12]. Ko, E., Soderstrom, M., Morgan, J. (2009). Development of perceptual sensitivity to extrinsic vowel duration in infants learning American English. J Acoust Soc Am. 2009 Nov; 126(5): EL134–EL139.

[13]. Kurihara, M., Chan, W. M., Chin, K. N., Martin-Lau, P., Nagami, M., Suthiwan, T., Suzuki, M. (2006). The Identification of Japanese Long and Short Vowels by Mandarin Chinese Speakers. CLaSIC 2006: Processes and Process-Orientation in Foreign Language Teaching and Learning. The 2nd CLS International Conference, CD-ROM. National University of Singapore, pp. 451-460.

[14]. Kubozono, H. (1995). Perceptual evidence for the mora in Japanese. In B. Connell and A. Arvaniti (eds.) Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV. (pp.141-156). Cambridge: Cambridge University Press.

[15]. Lawrence J. R. (1971). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. J. Acoust. Soc. Am. 51: 1296-1303.

[16]. Minagawa-Kawai, Y. (2001). Preciseness of temporal compensation in Japanese mora timing. Phonetica 2001, 58, 1-25.

[17]. Ong, J.H., Burnham, D., Escudero, P. (2015). Distributional Learning of Lexical Tones: A Comparison of Attended vs. Unattended Listening. PLoS ONE 10(7): e0133446.

[18]. Pająk, B., Levy, R. (2012) Distributional learning of L2 phonological categories by listeners with different language backgrounds. In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), Proceedings of the 36th Boston University Conference on Language Development (pp. 400-413). Somerville, MA: Cascadilla Press.

[19]. Rodgers, B., Purnell, T., Salmons, J. (2008). Expanding the set of acoustic features of the post‐vocalic voicing contrast in English. Journal of the Acoustical Society of America. 123, 3888.

[20]. Recasens, D. (2004). The effect of syllable position on consonant reduction

(evidence fromCatalan consonant clusters). Journal of Phonetics, 32 (2004), 435–453.

[21]. Saffran, J. R., Johnson, E. K., Aslin, R.N., Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. Cognition. 1999 Feb 1;70(1):27-52.

[22]. Wanrooij, K., Boersma, P., & Van Zuijen, T. L. (2014a). Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study. Frontiers in Psychology. Language Sciences, 5, Article 77, 1-12.

[23]. Wanrooij, K., Boersma, P., & Van Zuijen, T. L. (2014b). Distributional Vowel Training Is Less Effective for Adults than for Infants. A Study Using the Mismatch Response. PloS One 9, 1–13.

[24]. Xu, N., & Demuth, K. (2012). Acquisition of English codas by Mandarin-speaking adult learners of English. Proceedings of the 14th Australasian International Conference on Speech Science and Technology, pp. 41–44. Sydney: Causal.

[25]. Zhang, Y., Hayashi, R. (2015). Perception of syllable-final consonants by Chinese speakers and Japanese speakers, Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0389.