THE INFLUENCE OF TONAL LANGUAGE EXPERIENCE ON SPEECH PERCEPTION

Bachelor's Project Thesis

Xuechun Wu, s3283399, x.wu.17@student.rug.nl

Supervisors: Dr. Frank Tsiwah

Abstract: There are more than 7000 languages spoken in the world. Among these languages are tonal languages. They are common in East and Southeast Asia, and Africa. In tonal languages, a syllable usually carries a specific pitch. While those pitches convey information such as focus and emotion in non-tonal languages, in tonal languages, these pitches are used to convey different lexical and meanings. It is interesting to investigate if the language experience of tonal language supports people to distinguish tonal differences in unfamiliar tonal languages and non-linguistic tones. The hypothesis is that language experience and knowledge can help people distinguish tones in unfamiliar tonal languages and make accurate and quick judgments. A total of 20 native Mandarin speakers were recruited from the University of Groningen for this study. The results showed that the participants had high accuracy in judging Akan tones with a short response time.

1. Introduction

Every day, people receive and process external signals either actively or passively. The ability associated with it is called perception. Perception is a series of processes that organize and interpret the sensory information generated by external objects and events. Among them, speech perception refers to the listener's recognition and understanding of sounds, syllables, vocabulary, etc. That is, the process of establishing meaning from sounds. This is done through a series of processes. First of all, the brain needs to convert analog signals into different levels of digital signals (e.g. phonemes, syllables, and words). Secondly, while the brain remains sensitive to speech signals, it must also ensure cognitive flexibility, that is, the ability to adapt to different accents. Thirdly, the brain needs to segment sentences and find the space between words. Because there is no obvious pause between words when people speak, we subjectively need to separate them. Finally, speech signals require cognitive processing of grammatical vocabulary (Čistovič et al., 1968). These processes have to be carried out very quickly.

According to Ethnologue (2022), there are 7,151 languages spoken in the world. Among these languages are tonal languages. They are common in East and Southeast Asia, Africa, the Americas, and the Pacific. In tonal languages, a syllable usually carries a specific pitch. While those pitches convey information such as focus and emotion in non-tonal language, in tonal languages, these are used to convey different lexical meanings. Tonal and pitch-accent languages make up about 70% of the world's languages (Yip, 2002). In this article, we explore and discuss the impact of language experience on speech perception.

Burnham et al. (2015) examined both visual and auditory perceptual abilities, particularly tone perception, among people of different native languages. The native languages of the participants tested were Thai, Mandarin, Cantonese, Swedish and Australian English. Thai, Cantonese and Mandarin are tonal languages (pitch changes of a single syllable), Swedish is a pitch-accent language (relative pitch changes between successive syllables), and English is a non-tonal language. Experimental results show that visual information (such as lip, face and head motions) can enhance the listener's perception of tone and such visual augmentation is not affected by language experience. Regardless of whether the native language was a tonal language, a pitch-accent language or a non-tonal language, the effect of visual augmentation was equally observed. However, compared with tonal language and pitch-accent language perceivers, non-tonal language (English) perceivers use visual information more effectively for tone discrimination. This may be because tonal and pitch-accent language perceivers are more familiar and experienced with tone and therefore rely more on acoustic cues. But non-tonal language perceivers have no relevant experience, and they will try to obtain information from multiple sources to combine intonation discrimination (Burnham et al., 2015).

Moreover, in conditions where only auditory information is available, language experiences affect the ability of lexical tone perception. For example, English speakers' perception of speech pitch can be attenuated due to interference from the environment or other sounds, but tonal and pitch-accent perceivers were not disturbed. That is, lexical tone perception experience seems to contribute to the perception of non-native tones. However, the is not clear-cut. . Burmese is a tonal language that is cued by phonation type and duration. But Burmese listeners' accuracy in discriminating the tones of Mandarin words did not improve because their native language was a tonal language (Tsukada & Kondo, 2019). The perceptual abilities of Burmese listeners vary widely between different tone pairs. This may be because Burmese listeners try to substitute their native language when distinguishing tones, but because Burmese and Mandarin have different phonetic features, the substitution will only confuse them more. Burmese basically uses phonetic features other than fundamental frequency (F0; pitch), so Burmese listeners may tend to ignore the change in F0, which is necessary to distinguish the tones of Mandarin words. But compared to non-tonal language listeners, Burmese listeners may be able to learn Mandarin more easily because of relevant experience (Cooper & Wang, 2012).

However, some studies have also shown that different language backgrounds do not significantly help the learning of tonal languages (Tsukada, Xu & Rattanasone, 2015). They compared the discriminative accuracy of Mandarin lexical tones among four groups of participants with different language backgrounds (Mandarin native speakers, Cantonese native speakers-Mandarin learners, English native speakers-Mandarin learners, and native English speakers) and found that there was no significant difference in the accuracy of the two learner groups in distinguishing Mandarin tones. That is to say, the experience of one tonal language does not promote nor inhibit the learning of another tonal language. Experience with tonal language may be used for identification tasks rather than discrimination tasks (Tsukada, 2011).

Hao's (2018) study investigated whether English speakers have difficulty distinguishing Mandarin tones due to a lack of experience with native tones. The experiment compared three groups of English speakers with different levels of Mandarin to see whether increasing Mandarin experience would alter their perceived sensitivity to Mandarin tones. The result shows that the experience of Mandarin affects native English speakers' perception of Mandarin tones. Without training, native English speakers are less sensitive to pitch changes. Because the pitch of a single syllable is not essential for distinguishing lexical meanings in English (Hao, 2018). Native Korean speakers who are learning Mandarin performed similarly. Listeners have a relatively low resolution for a specific tone pair (e.g. Tone 2 - Tone 3 in Mandarin), but with the increase in Mandarin experience, the listener's perception of this tone pair becomes more acute (Tsukada & Han, 2019). Native Japanese speakers also have a weak perception of the Tone 2 - Tone 3 tone pair in Mandarin, regardless of gender. Consistent with the experimental results of Burnham et al. (2015), the perceptual abilities of listeners experienced in tonal language are less affected by environment and gender (Tsukada, Kondo & Sunaoka, 2016).

Mandarin is a typical tonal language and the syllable structure of Mandarin is very simple. A syllable has four phonemes at most, and there are no consecutive consonants (such as 'lighting' in English). There are four tones represented by diacritics in Mandarin. Tone 1: a high-level tone, represented by a macron (m ā). Tone 2: rising tone, represented by an acute accent (má). Tone 3: slight fall followed by a rising tone, represented by a caron (mǎ). Tone 4: a falling sound, represented by a grave accent (mà). Mandarin also has neutral light tones, whose pitch is mainly determined by the tone of the preceding syllable.

Similarly, Akan is also a tonal language. Akan is the main native language of the Akan people of Ghana and is spoken in most of southern Ghana. Akan can be roughly divided into three dialects: Fante, Asante Twi and Akwapim Twi. These three dialects can communicate with each other, Asante Twi is the most widely used. Akan belongs to the Kwa group of the Niger-Congo language family. Tone, vowel harmony and nasalization are very specific linguistic features of Akan. There are two phonemic tones in Akan, high /H/ (an acute accent) and low /L/ (a grave accent) (Schachter & Fromkin, 1968).

In different tonal languages, such as Akan and Mandarin, the types and patterns of tones are also different. Therefore, the aim of this study is to investigate can a native speaker of a tonal language transfer their knowledge or ability to distinguish between minimal pairs to another tonal language to which they have no exposure? Also, can native Mandarin speakers be able to accurately distinguish the tonal difference between two syllables? According to existing research, lexical tone perception is influenced by prior language experience, so the hypothesis is that language experience and knowledge can help people distinguish tones in unfamiliar languages, and make accurate and quick judgments. An experiment has been designed to explore these questions and verify the assumptions mentioned above.

For this study, 20 native speakers of Mandarin were recruited from the University of Groningen. The stimulus for the experiment was 120 minimal pairs. These pairs consist of the same number of words in three languages (Akan, Mandarin and pure tones). Each pair contains two words that are the same or different. During the experiment, participants listened to stimuli using headphones and responded by pressing buttons on the keyboard. They were asked to judge whether the two words in each pair were the same or different.

2. Method

2.1. Participants

A total of 20 participants (7 males and 13 females) were recruited from the University of Groningen for this study. The age range of the participants was from 21 to 31, with a mean age of 22.55. All of them were native Mandarin speakers. None of the participants had any language experience or knowledge that is related to Akan. According to surveys before the experiment, none of the participants had hearing impairments, and all were right-handed. All participants were tested individually in a quiet room on the University of Groningen campus. The project was approved by the Research Ethical Review Committee of the Faculties of Arts, Philosophy, and Theology and Religious Studies of the University of Groningen, and participants were paid for their participation. Also, all participants signed an informed consent before participating in the experiment.

2.2. Stimuli

The same number of Akan and Mandarin mono-syllabic words and pure tones made up a total of 120 minimal pairs. So there are 40 pairs for each type of language, half of which are the same pair (such as shǒu-shǒu), and the rest are different (such as bái-bǎi) for Mandarin, and pám-pám (the same condition) and dàà- dá (different condition) for Akan. Akan and Mandarin words were recorded in noise-free studios by corresponding male native speakers. Before the recording, the researchers confirmed that they knew the exact pronunciation of the words. During the recording, each word on the reading sheet was read twice. The recording was supervised by another native speaker to ensure accuracy. The pure tones were created by using audacity software.

2.3. Procedure

The task required participants to press keys to judge the minimal pairs they heard. Participants listened to stimuli via a Beats headphone connected to a Dell laptop in a studio on campus. Participants judged the two mono-syllabic words in each pair, pressing the “ Q ” key on the keyboard when they thought the two mono-syllabic words were the same, and the “P” key when they thought the two mono-syllabic words were different. The next pair was played only after the participant had responded. Reaction time is not limited, but participants were asked to respond as quickly as possible without sacrificing accuracy. All participants listened to the same 120 pairs, but the pairs were played in random order. All pairs were divided into two blocks, each block had 60 pairs, and there was a 5-minute rest period between the two blocks. This is to prevent participants from being unable to concentrate due to repeating similar operations for a long time. Because inattention can lead to unnecessary errors in judgment.

Participants read the instructions for the experimental procedure prior to the start of the experiment, and all questions were explained by the researcher. Before the formal experiment began, each participant had the opportunity to practice (the practice contains 24 pairs and 8 pairs for each type). This is to familiarize the participants with the experimental procedure and to ensure that they all understand the keyboard operation correctly. Experimental data for warm-up practice were not recorded and were therefore not included in the data analyses. In addition, the participants did not obtain any information about the specific content of the experiment to be tested in advance. That is, their judgments were not affected by deliberate attention to the tones in the mono-syllabic words.

2.4. Statistical analysis

Response time (RT) is measured from the end of the pair presentation. Longer RT means higher perceived difficulty. Trials with response times less than or equal to 100ms and above 3000ms were considered to be outliers, and thus, were removed from the data. Because the sample data may not be truly independent, data analysis used linear mixed models and the lme4 package (Douglas, Martin, Ben & Steve, 2015) in R. The model was built to investigate how language affects response time, so the dependent variable in the model was RT and the fixed effect was Language. In the original model, as random effect, there is an intercept for subject. Because different subjects may be distributed differently, and different subjects may be affected by Language differently, random intercept for subject was also added to the model. To ensure normal distribution, response time data were logarithmically transformed.

3. Result

3.1. Accuracy rate

Figure 3.1 shows the accuracy of participants' responses to stimuli in different languages. After removing outliers, there are 2237 valid data (Akan: 727, Mandarin: 753, Pure Tone: 757). The "0" in the figure (the red bar) indicates the

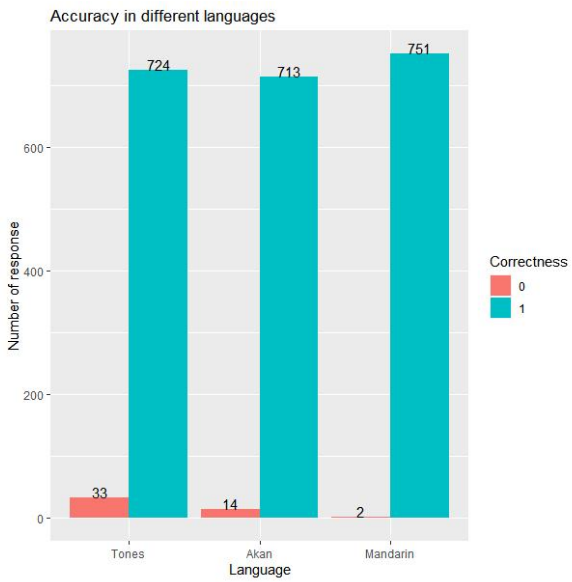

Figure 3.1: Graph of accuracy in different languages. Correctness 0 refers to incorrect response and correctness 1 refers to correct response.

incorrect response, and the "1" (the cyan bar) indicates the correct response. The overall accuracy was 97.8%. Mandarin had the highest accuracy at 99.7%, with only 2 incorrect responses. Next is Akan with 98.1% accuracy. Participants performed the worst on the Tones pairs, with 33 incorrect responses and an accuracy of 95.6%.

3.2. Response times

Figure 3.2 shows participants' response times to different languages. Obviously, Akan has the

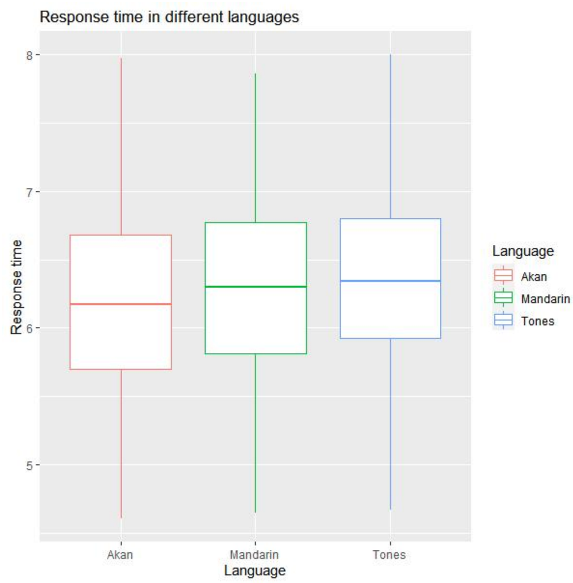

Figure 3.2: Graph of response time in different languages.

shortest response time, followed by Mandarin. Tones had the longest response time. A shorter response time means lower perceived difficulty. In order to determine if they were significantly different, a linear mixed model was built using the lme4 package (Douglas, Martin, Ben & Steve, 2015) in R. As mentioned in the Method section, the dependent variable in the model was response time, the fixed effect was Language, and the random effects were subject (with random slope Language and random intercept 0). The results showed that there was no significant difference between the response times of Akan (M = 6.16, SD = .08) and Mandarin (M = 6.26, SD = .05), $t(40)$ = 1.86, p = .07. But on the contrary, there was a significant difference between the response times of Akan (M = 6.16, SD = .08) and Tones (M = 6.36, SD = .06), $t(32)$ = 3.27, p = .003. There was also a significant difference between the response times of Tones (M = 6.36, SD = .06) and Mandarin (M = 6.26, SD = .05), $t(87)$ = -2.44, p = .02.

Figure 3.3 shows the response time of participants of different genders. The average response time of female participants (672ms) was slightly shorter than that of male participants (676ms). Among them, female participants responded faster to Tones and Mandarin, while male participants responded faster to Akan. However, the results of the analysis of variance (ANOVA) showed that gender had no significant effect on response time, p > .05. We used an alpha level of .05 for all statistical tests.

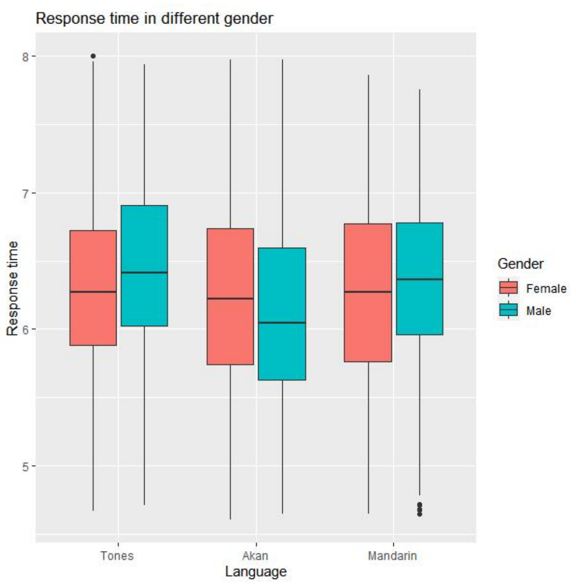

Figure 3.3: Graph of response time in different gender.

4. Discussion and conclusion

This study investigated whether native Mandarin speakers could accurately distinguish the tones in Akan, pure tones and Mandarin, and compared participants' response times to the different languages. The results clearly demonstrated that the participants, as native speakers of Mandarin, had high accuracy (98.1%) in judging Akan tones with a short response time. Although the participants made some mistakes in their judgment of Tones, the accuracy rate was still as high as 95.6%. So it can be said that the results support the view of Tsukada & Kondo (2019) that language experience can help perceive other languages. This is in line with the original hypothesis that language experience and knowledge can help people distinguish tones in unfamiliar languages, and make accurate and quick judgments. This also answers the research question, a native speaker of a tonal language can transfer their knowledge or ability to distinguish between minimal pairs to another tonal language to which they have no exposure. After comparing the participants' response times to the three languages, it was found that Akan's response time was the shortest. It is even shorter than the response time in the participants' native language. In post-experiment survey interviews, we learned that a significant proportion of the participants indicated that when they heard Mandarin words in the experiment, they unconsciously thought about the meaning of the words. Over-familiar words caused them to make subconscious associations. This is most likely the reason for their slower response to Mandarin words. Apart from that, Akan has simpler tones compared to the tones in Mandarin. There is only a single rising or falling key in Akan. But in Mandarin, there is a tone that has a slight fall followed by a rising tone. The changing tones may increase the difficulty of perception, resulting in a longer response time. It is worth mentioning that the response time of Akan and Mandarin is not significantly different, so it still fits our hypothesis.

However, surprisingly, participants took longer to react to the tones than to Akan and Mandarin. There is a significant difference between them. This means that, for the participants, the perceived difficulty of Tones is higher than that of Mandarin and Akan. We suspect this may be due to the frequency of the sound. Because the recordings of Akan and Mandarin we used in the experiment were recorded by adult males. Usually adult male voice frequency range is about 80Hz to 180Hz. This is well below the lowest frequency (327Hz) of the tones used in the experiment. However, there is no evidence that reaction time is affected by sound frequency. So this interpretation must be taken with caution.

Besides, gender did not have a significant effect on response time, that is, there were no gender differences in the perception of tones.

5. Future research

In this study, the question of why tones require longer reaction times was not answered. While it was mentioned above that it may be related to sound frequency, there is no evidence to support this speculation. So, this can be future research. By examining participants' reaction times between sounds of different frequencies, it was possible to investigate whether sound frequency affected response times.

In addition, this study investigated the perception of Akan tones by native Mandarin speakers. The tone of Akan is simpler than that of Mandarin, so the participants had a very good performance. However, would the same result be true for more complex tonal languages than Mandarin? For example, to investigate the perception of Cantonese tones by native Mandarin speakers. Cantonese has 9 tones, which is much more complicated than Mandarin, so the difficulty of perception will also increase. Investigating whether participants can perform as well as in this study can also be future research.

References

Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., ... & Webster, D. (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics, 36*(6), 1459-1491.

Čistovič, L., Golusina, A., Lublinskaja, V., Malinnikova, T ., & Žukova, M . (1968). Psychological methods in speech perception research. *STUF-Language Typology and Universals, 21*(1-6), 33-39.

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *The Journal of the Acoustical Society of America, 131*(6), 4756-4769.

Douglas, B., Martin, M., Ben, B., & Steve, W. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Gregory, R. L., & Zangwill, O. L. (Eds.). (1987). *The Oxford companion to the mind* (pp. 598-601). Oxford university press.

Hao, Y. C. (2018). Second language perception of Mandarin vowels and tones. *Language and Speech, 61*(1), 135-152.

Schachter, P., & Fromkin, V. (1968). *A Phonology of Akan: Akuapem, Asante, Fante* (pp. 105-107).

Tsukada, K. (2011). The perception of Arabic and Japanese short and long vowels by native speakers of Arabic, Japanese, and Persian. *The Journal of the Acoustical Society of America, 129*(2), 989-998.

Tsukada, K., & Han, J. I. (2019). The perception of Mandarin lexical tones by native Korean speakers differing in their experience with Mandarin. *Second Language Research, 35*(3), 305-318.

Tsukada, K., & Kondo, M. (2019). The perception of Mandarin lexical tones by native speakers of Burmese. *Language and speech, 62*(4), 625-640.

Tsukada, K., Kondo, M., & Sunaoka, K. (2016). The perception of Mandarin lexical tones by native Japanese adult listeners with and without Mandarin learning experience. *Journal of second language pronunciation, 2*(2), 225-252.

Tsukada, K., Xu, H. L., & Rattanasone, N. X. (2015). The perception of Mandarin lexical tones by listeners from different linguistic backgrounds. *Chinese as a second language research, 4*(2), 141-161.

Yip, M. J. W. (2002). *Tone* (Chap. 1, pp. 1-14). New York: Cambridge University Press.