university of groningen / faculty of science and engineering

The influence of language- and musical experience on speech perception

Bachelor's Project Thesis

Maya Swierstra, s3649261, m.h.swierstra@student.rug.nl
Supervisor: Dr F. Tsiwah

Abstract: Speech sounds come in through the ears and are then interpreted and understood in the brain. Speech perception is a field of study that researches how humans perceive, recognise and understand speech. Language experience during the first year of life has a big effect on perception of speech. One of the differences in language experience in this first year can be the difference between tonal languages and non-tonal languages. Tonal languages are languages in which pitch is used to convey a difference in meaning between words. Previous research has shown that experience with a tonal language provides an advantage for distinguishing between lexical tones in another tonal language. Additionally, in an experiment about tonal language in combination with music, it has been found that experience with a tonal language decreases performance in determining differences in pitch. The current research investigates the performance in distinguishing between sounds when native Dutch speakers are presented with minimal pairs in Dutch, pseudowords, Akan and pure tones. Additionally, an analysis is performed taking into account the musical background of the participants. It is found that response times are significantly higher for words in Akan than for Dutch words, pseudowords and pure tones. Additionally, no significant difference in response times was found between participants with a musical background and participants with no musical background.

1 Introduction

The language experience that infants get during the first year of life has a big effect on perception and production of speech (Kuhl, 1998). For example, children who grow up in America will get a very different overall language experience than children who grow up in Japan, and therefore they will develop a different manner of perception and production of speech. An example of this has been discussed by Kuhl et al. (2006). In this experiment, 6-8 and 10-12 month old infants from America and Japan were examined on developmental change using the /r-l/ phonemic contrast. When the infants were tested at 6-8 months old, the performance of distinguishing between the /r/ and /l/ sounds were similar between the American- and Japanese group. However, when they were tested again when the infants were between 10-12 months old and therefore had had more experience in and exposure to their native language, the American infants performed significantly better, while the Japanese infants performed significantly worse than before. This is one of many results that support the hypothesis that an adult's knowledge of a language is a product of many years of exposure to a specific language environment (Strange & Jenkins, 1978). Apart from contrasts between for example /r/ and /l/, another contrast in language environment can be tonal- and non-tonal languages.

Although Indo-European languages such as English use pitch to convey information about prosody, stress and emotional content, they are not considered to be tonal languages. In tonal languages, pitch is not only used for prosody, stress and emotional content, but most importantly, it is used to convey difference in meaning between words. Examples of tonal languages are Chinese, Vietnamese and Thai (Burnham et al., 2015). Unlike non-tonal languages which have two phonological features, namely consonants and vowels, tonal languages have three phonological features: consonants, vowels and lexical tone. By using different tonal categories on the same syllable, differences in meaning can be conveyed. An example of this is the Mandarin syllable /ya/. When using Tone 1 on the syllable, it means duck. When using Tone 3 however, /ya/ means elegant (Yu et al., 2022).

Another tonal language is Akan, which is a sub-Saharan African language most widely spoken in Ghana. About 80 percent of the population speak Akan as their first or second language. Akan is also spoken in the central and eastern part of Cote d'Ivoire. Akan, as a tonal language, has two basic level tones: high and low tone. The high and low tones are used for both lexical and grammatical functions (Dolphyne, 1988). Lexical high tones are transcribed as [x́], while lexical low tones are transcribed as [x̀]. An example of a word that differs in the lexical sense when using different tones is kooko: /kóókó/, with two high tones, means porridge. /kòòkó/, with first a low and then a high tone, means hemorrhoid. Grammatical differences can also be established using different tones. An example is the sentence /Yaw gyíná hɔ/, which means Yaw stands there. /Yaw gyìnà hɔ/ , however, means Yaw is standing there. The only difference between these two sentences is the lexical tone of gyina, but in the first sentence the verb has the habitual form and in the second sentence the verb has the stative form (Dolphyne, 1988).

Several experiments have been conducted on tonal language perception. An example is an experiment conducted by Burnham et al. (2015), in which it was investigated how a difference in linguistic experience might influence the mechanisms of linguistic tone perception as compared to non-linguistic tone perception. In this experiment, native Thai, native Cantonese, native Swedish (a pitch-accent language, which is a subcategory of tonal languages) and native English (naïve to tonal language) participants performed three AX discrimination tasks. These are tasks in which the subject has to determine whether the two presented stimuli are the same or different. Auditory Thai tone contrasts were modified into two non-speech formats, namely low pass filtered speech and violin sounds, while keeping the pitch, or F0, constant. The three AX discrimination tasks that were performed were identical except for the stimulus format; there was one task with normal speech, one task with low pass filtered speech and one task with violin sounds. Participants were required to listen to pairs of stimuli and indicate whether the stimuli they heard were different or the same. It was hypothesized that native English speakers would have an increased performance in discriminating the same F0 patterns when the presented stimuli are in a non speech context as opposed to a speech context. Additionally, it was hypothesized that for the native Thai, Cantonese and Swedish speakers, there would be no difference in performance between the non speech context and the speech context. The results showed that for native English speakers naïve to tonal language, performance indeed improved for the non speech context condition as opposed to the speech context condition. It also confirmed that there was close to no difference across speech, low pass filtered speech and violin sounds for the native Thai, Cantonese and Swedish speakers. Overall, it was found that experience with a tonal- or pitch-accent language provides an advantage for distinguishing between lexical tones.

Another study on tonal language perception has been conducted by Tsukada & Kondo (2019). This experiment tested native speakers of Burmese on the perception of Mandarin. Burmese is a tonal language of which the tones are cued by pitch as well as phonation type, whereas Mandarin is a language of which the tones are cued primarily by pitch. It was hypothesized that native speakers of Burmese would be able to distinguish between unfamiliar Mandarin sounds better than native speakers of non-tonal languages, like English. This hypothesis was based on the prediction that Burmese speakers would be facilitated by their knowledge of their first language. There were three participant groups; Australian English, Burmese and Mandarin listeners. Tone discrimination accuracy was measured by conducting the four-alternative forced-choice discrimination task. In this task, participants listened to three monosyllabic words per trial, differing in lexical tones. The participants were asked to determine whether there was an odd one out between the three words. Like this, the accuracy of perception of six tone pairs was assessed. It was found that Mandarin listeners had a higher accuracy than non-native listeners in distinguishing between tone pairs. Additionally, Australian English listeners had a higher accuracy than Burmese listeners in discriminating nearly all tone pairs. Overall, the results suggested that knowledge of lexical tones in one's first language might not be advantageous in the perception of non-native tones. Additionally, the use of phonation type by Burmese listeners might have added to the less than optimal performance in distinguishing between Mandarin tones.

In addition to experiments on tonal language alone, research has been performed on tonal and non-tonal language in combination with musical experience. In an experiment by Stevens et al. (2011), it was hypothesized that tonal language speakers are more sensitive to pitch change and pitch contour than non-tonal language speakers. Three tasks were performed; firstly it was examined whether a tonal language background influenced the discrimination of contour in linguistic items. Secondly it was investigated whether tonal language background affects the discrimination of pitch change in music intervals. Lastly a frequency discrimination task was performed in which it was investigated whether tonal language speakers have a lower difference threshold for tone frequency discrimination than non-tonal language speakers. To realise this, participants were asked to perform AX discrimination tasks. The results suggested that the tonal language speakers had a significantly increased performance of speed and accuracy in distinguishing between speech items based on pitch contour. It was also found that the tonal language speakers were significantly faster in discriminating pitch change in music intervals, while the accuracy was equivalent across the tonal language group and the non-tonal language group.

Another experiment has been conducted on absolute pitch in combination with musical expertise and tonal language background (Van Hedger & Nusbaum, 2018). It was found that participants who are tonal language speakers had a decreased accuracy in determining pitch. Additionally, tonal language speakers showed less musical expertise than non-tonal language speakers.

A phenomenon that can be of importance when researching (tonal) language perception is the recognition of pseudowords, as these are used as stimuli in a great number of psycholinguistic experiments. Kelley & Tucker (2022) examined how pseudowords are recognised. The hypothesis was that the processing of pseudowords uses the same architecture as real words. An auditory lexical decision task was performed in which participants had to classify words and pseudowords. The results suggested that the processing of pseudowords indeed involves the same processing mechanisms as those of real words.

The current study tests the hypothesis that the speed and accuracy of distinguishing between speech sounds by native Dutch speakers who are naïve to tonal language, improve when presented words in Dutch as opposed to when presented words in Akan, an unknown tonal language. The study also tests the hypothesis that the speed and accuracy of judging tonal differences by native Dutch speakers increases in a non-linguistic context (i.e. distinguishing between tones) as opposed to a linguistic context (i.e. distinguishing between words in Akan). Another question that will be addressed is that of whether the speed and accuracy of distinguishing between speech sounds by native Dutch speakers change when the presented words have a meaning in Dutch as opposed to when the presented words are pseudowords. Additionally, it will be investigated whether participants with a musical background will have an increased or decreased performance in distinguishing between presented stimuli as opposed to participants without a musical background.

To examine these hypotheses and questions, an AX discrimination task including Dutch words, pseudowords, Akan words and tones will be performed by native Dutch speakers.

2 Method

2.1 Participants

A total of twenty native Dutch speakers were tested (8 males and 12 females; M = 22.7 years, SD = 2.1, range 20 - 28). The participants were naive to tonal language. 13 participants reported having a musical background. One participant reported to be left handed. Participants were compensated for participating in the experiment and also signed a form of informed consent prior to the experiment. 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.

2.2 Materials

Stimulus sets of four different types were created: Dutch words, pseudowords, Akan words and tones. Lists of minimal pairs were created in Dutch, Akan and pseudowords. The minimal pairs in Dutch and the pseudowords consisted of one syllable words that differed only in the first consonant. The words were recorded one by one by a native Dutch speaker. Similarly, the minimal pairs in Akan consisted of one syllable words that differed only in lexical tone. These words were recorded one by one by a native Akan speaker. The recordings in each language were then combined into pairs in such a way that per stimulus type, 15 pairs were the same, and 15 pairs were different. Two additional pairs per stimulus type were created for the practice phase. There was a gap of 250 ms between the two words in each pair.

For the stimulus type Tone, 30 tonal pairs were created using Audacity software. 15 of these tonal pairs were the same and 15 pairs were different. For the pairs that were different there were five conditions: The pairs could be 67 percent, 75 percent, 85 percent, 90 percent or 95 percent similar. Three pairs of each similarity were created. An example of the 95 percent similarity condition pair is one tone of 342 Hz and one tone of 360 Hz. The tones ranged from 342 Hz to 1893 Hz. The individual tones were 750 ms long each and there was a gap of 250 ms between the two tones in each pair.

A total of 136 stimuli was created for the whole experiment, 16 stimuli for the practice phase and 120 stimuli for the main experiment. See Table 2.1 for a sample of the used stimuli. See appendix A for the complete list of used stimuli.

Table 2.1: A sample of the used stimuli

Dutch	
Same	**Different**
tas tas (bag)	wit zit (white sit)
vis vis (fish)	gat kat (hole cat)
Akan	
Same	**Different**
bɔ́ bɔ́	wɔ̀ wɔ́
twí twí	nyà nyá
Pseudowords	
Same	**Different**
dos dos	ris tis
mef mef	sil zil
Tone	
Same	**Different**
396Hz 396Hz	611Hz 458Hz
440Hz 440Hz	951Hz 637Hz

2.3 Procedure

All experiments were conducted using the same noise cancelling headphones (Bose Quietcomfort 35 wireless headphones II) and the same laptop. Before the experiment, the participants received a written explanation of what they were about to do. Each participant also signed a form of informed consent.

Participants completed an AX discrimination task that was created and set up in OpenSesame software (Mathôt et al., 2012). The task started with an in depth explanation of what was expected of the participant. The experiment continued with a practice phase in which participants were familiarised with the format of the experiment. The participants were required to concentrate on a white fixation dot on the screen while listening to the stimulus. After the stimulus was done playing, the dot would turn red, after which the participant could enter their response, namely the ‘p’ key if the two words or tones were the same, and the ‘q’ key if the two words or tones were different. The participants were asked to do this as quickly and accurately as possible. After entering the response, the next stimulus started playing. The practice phase consisted of 16 trials. After the practice phase the participants could continue with the real experiment which consisted of 120 trials with a mandatory three minute break after 60 trials. All trials were randomised for each participant to omit the chance of fatigue influencing the results. After finishing the experiment, participants were debriefed.

2.4 Data Analysis

The collected data was preprocessed in R (R Core Team, 2022) prior to conducting statistical analysis. Firstly the data was log transformed to create a better normally distributed fit data set for analysis. Outliers were removed by excluding response times of less than 100 ms or more than 2500 ms from the data. The data of the left handed participant was excluded as well. Data of both correct and incorrect responses were used.

A linear mixed effects analysis of the relationship between reaction times and stimulus type (Dutch, pseudowords, Akan, tones) was performed using the lmer function of the lme4 package (Bates et al., 2015) in R (R Core Team, 2022). Because it was important to account for variation between participants and items, linear mixed effects regression modelling was used. Several models were created using reaction time as dependent variable. Intercepts for subjects and items were included as random effects. Stimulus type, musical background and gender were added as fixed effects, as well as the interactions between these variables. Based on the AIC and the log likelihood ratio tests of the full model with the effect in question against the model without the effect in question, the best model was chosen.

3 Results

3.1 Reaction times

The log reaction times from the AX discrimination task per stimulus type can be seen in Figure 3.1.

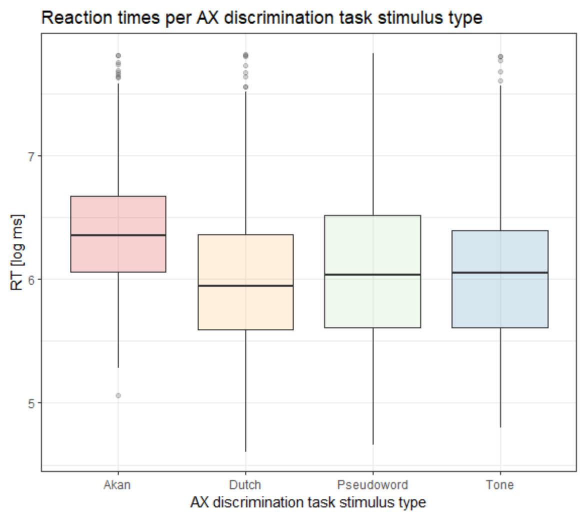

Figure 3.1: The log reaction times from the AX discrimination task per stimulus type.

The reaction times for Akan stimuli is higher than those for the other three stimulus types. There is a negligible difference between the reaction times for Pseudoword and Tone stimuli, while the reaction times for Dutch stimuli are the lowest of the four.

In Figure 3.2, the log reaction times from the AX discrimination task per stimulus type are shown, distinguishing between participants with a musical background and participants with no musical background. The reaction times are generally higher for people with a musical background than for people with no musical background. The biggest difference in reaction times between the two groups is for stimulus type Tone, while the smallest difference in reaction times between the two groups is for stimulus type Dutch.

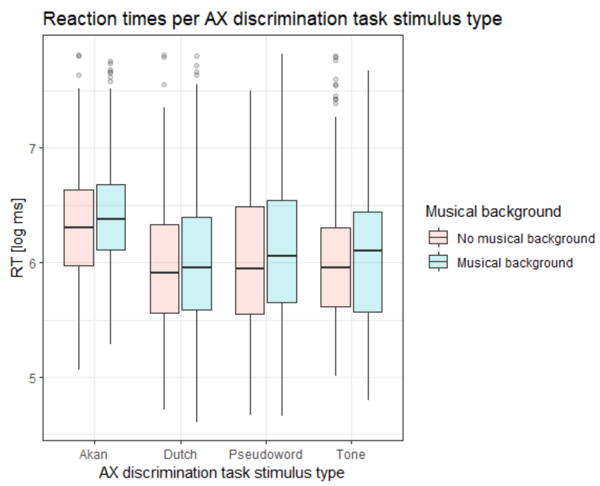

Figure 3.2: The log reaction times from the AX discrimination task per stimulus type, taking into account musical background.

Figure 3.3 depicts the log reaction times from the AX discrimination task per stimulus type taking into account gender of the participants. For all stimulus types, female participants generally have higher reaction times than male participants.

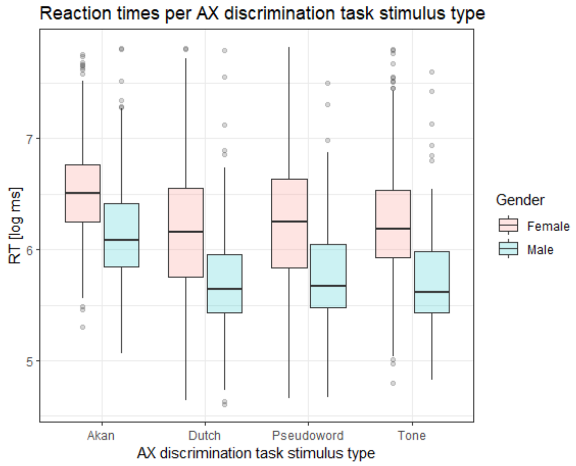

Figure 3.3: The log reaction times from the AX discrimination task per stimulus type, taking into account gender.

3.2 Accuracies

There did not seem to be any major differences in the accuracies from the AX discrimination task across the different stimulus types. See appendix B for the graphs of the results of the accuracies.

3.3 Statistical analysis

The initial complete model included reaction time as dependent variable, intercepts for subjects and items as random effects, and stimulus type, musical background and gender as fixed effects. The model was constructed in R as follows:

```
model.null <- lmer(reaction_time ~
stimulus_type + musical_background
+ gender + (1 | subject) + (1 |
sound_file), data)
```

See Table 3.1 for the output of the initial complete model. According to the model, stimulus type does seem to influence reaction time significantly (p<0.001) and so does gender (p<0.01). However, musical background does not seem to influence reaction time significantly (p = 0.892). Therefore, musical background was not taken into account in further analysis.

Stimulus type influenced reaction time significantly (output of the model without stimulus type as fixed effect versus the model with stimulus type as fixed effect: $X^2(3) = 69.23, p < 0.001$, with a lower AIC value for the model with stimulus type as fixed effect).

Gender influenced reaction time significantly (output of the model without gender as fixed effect versus the model with gender as fixed effect: $X^2(1) = 8.65, p < 0.01$, with a lower AIC value for the model with gender as fixed effect).

There was a significant interaction between Stimulus type and Gender (output of the model without interaction versus the model with interaction: $X^2(3) = 11.17, p < 0.05$, with a lower AIC value for the model with interaction between Stimulus type and Gender).

Based on these results, the best model is defined as follows:

```
model <- lmer(reaction_time ~ stimulus_
type * gender + (1 |subject) + (1 |
sound_file), data)
```

See Table 3.2 for the output of the final complete model.

The reaction times for stimulus type Dutch were significantly lower than the reaction times for stimulus type Akan (β = -0.37, SE = 0.05, t = -7.49, p <0.001), and so were the reaction times for stimulus type Pseudoword (β = -0.28, SE = 0.05, t = -5.78, p <0.001) and Tone (β = -0.28, SE = 0.05, t = -5.64, p <0.001).

The reaction times of male participants were significantly lower than the reaction times of female participants (β = -0.36, SE = 0.14, t = -2.58, p <0.05).

There was a statistically significant interaction between gender and stimulus type Dutch (β = -0.09, SE = 0.05, t = -1.77, p <0.1), between gender and stimulus type Pseudoword (β = -0.11, SE = 0.05, t = -2.31, p <0.05) and between gender and stimulus type Tone (β = -0.16, SE = 0.05, t = -3.25, p <0.01).

4 Conclusion and discussion

The current study attempted to research the influence of language- and musical experience on speech perception. This was done by conducting an AX discrimination task on native Dutch speakers in-

Table 3.1: Initial complete model on native Dutch speakers' reaction times

	Estimate	Std. Error	df	t value	Pr(>| t |)
(Intercept)	6.54290	0.12680	21.46174	51.599	<2e-16 ***
stimulus_typedutch	-0.40122	0.04570	118.04331	-8.780	1.51e-14 ***
stimulus_typepseudo	-0.32564	0.04558	116.86167	-7.144	8.35e-11 ***
stimulus_typetone	-0.33597	0.04554	116.48428	-7.377	2.58e-11 ***
response_musicaly	0.01883	0.13665	18.98695	0.138	0.89186
response_genderm	-0.44858	0.13665	18.98714	-3.283	0.00392 **

Table 3.2: Linear mixed model output on native Dutch speakers' reaction times

	Estimate	Std. Error	df	t value	Pr(>| t |)
(Intercept)	6.52208	0.08895	25.38660	73.327	<2e-16 ***
stimulus_typedutch	-0.36840	0.04921	158.52763	-7.486	4.62e-12 ***
stimulus_typepseudo	-0.28353	0.04908	156.89179	-5.778	3.96e-08 ***
stimulus_typetone	-0.27669	0.04906	156.72644	-5.640	7.76e-08 ***
response_genderm	-0.35989	0.13939	20.87020	-2.582	0.01744 *
stimulus_typedutch:response_genderm	-0.08899	0.05015	2059.70433	-1.774	0.07614 .
stimulus_typepseudo:response_genderm	-0.11443	0.04963	2057.79765	-2.306	0.02121 *
stimulus_typetone:response_genderm	-0.16051	0.04942	2058.05838	-3.248	0.00118 **

cluding Dutch words, pseudowords, Akan words and tones.

The results from the conducted experiment can answer the earlier posed research questions. Firstly, the results support the hypothesis that the speed of distinguishing between speech sounds by native Dutch speakers who are naïve to tonal language improves when presented words that were in Dutch as opposed to Akan, an unknown tonal language. Secondly, the results support the hypothesis that the speed of judging tonal differences by native Dutch speakers increases in a non-linguistic context as opposed to a linguistic context.

Additionally it has been found that native Dutch speakers have a higher reaction time for distinguishing between pseudowords than Dutch words. Therefore, these results support the hypothesis that semantics facilitate speech perception. The reaction times of pseudowords and tones were almost identical, but both significantly lower than the reaction times to Akan stimuli. This is because pseudowords and tones are both devoid of meaning, but both are in a way familiar to Dutch people.

Finally it was shown that native Dutch speakers with a musical background do not have an increased performance for distinguishing between presented stimuli than native Dutch speakers with a non musical background. The results in Figure 3.2 even suggest that this is the other way around; a pattern was found where native Dutch speakers with a musical background had a higher reaction time than native Dutch speakers with a non musical background. However, since a significant difference has not been found, further research must be conducted to support or reject this suggestion.

An interesting result is that gender significantly influenced reaction time. This was not hypothesized but does support previously conducted research by for example Blough & Slavin (1987), Adam (1999) and Karia et al. (2012). These experiments concluded that there is an overall reaction time advantage for male participants, possibly suggesting differences in processing strategy between male- and female participants. This might be caused by a difference in strength of motor response between male- and female participants (Jain et al., 2015).

For all research questions it was established that there was no significant difference in accuracy. This might suggest that the overall experiment was not challenging enough. Increasing the difficulty of the experiment might result in finding significant differences in accuracy as well as reaction times of participants with a musical- and non-musical background, where the current experiment did not. This could be done by for example creating and using stimuli pairs that are more difficult to distinguish from each other. All in all, the results show an effect of language experience on speech perception. Increasing difficulty of the experiment might show said effect even clearer. It might also show an effect of musical experience on speech perception.

References

Adam, J. (1999). Gender differences in choice reaction time: evidence for differential strategies. *Ergonomics*, *42*(2), 327-335. doi: https://doi.org/10.1080/001401399185685

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Blough, P., & Slavin, L. (1987). Reaction time assessments of gender differences in visual-spatial performance. *Perception Psychophysics*, *41*, 276-281. doi: https://doi.org/10.3758/BF03208225

Burnham, D., Kasisopa, B., & Reid, A. (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics*, *36*, 1459-1491.

Dolphyne, F. A. (1988). *The akan (twi-fante) language: Its sound systems and tonal structure.* Accra: Universities of Ghana Press.

Jain, A., Bansal, R., Kumar, A., & Singh, K. D. (2015). A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *International journal of applied basic medical research*, *5*(2), 124–127. doi: https://doi.org/10.4103/2229-516X.157168

Karia, R., Ghuntla, T., Mehta, H., Gokhale, P., & Shah, C. (2012). Effect of gender difference on visual reaction time: A study on medical students of bhavnagar region. *IOSR Journal of Pharmacy*, *2*(3), 452-454.

Kelley, M. C., & Tucker, B. V. (2022). The recognition of spoken pseudowords. *Language, Cognition and Neuroscience*. doi: 10.1080/23273798.2022.2053729

Kuhl, P. (1998). Effects of language experience on speech perception. *The Journal of the Acoustical Society of America*, *103*(5), 2931. doi: https://doi.org/10.1121/1.422159

Kuhl, P., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2). doi: https://doi.org/10.1111/j.1467-7687.2006.00468.x

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324. doi: 10.3758/s13428-011-0168-7

R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Stevens, C. J., Keller, E. P., & Tyler, M. D. (2011). Tonal language background and detecting pitch contour in spoken and musical items. *Psychology of Music*, *41(1)*, 59-74.

Strange, W., & Jenkins, J. (1978). Role of linguistic experience in the perception of speech. *Perception and Experience. Perception and Perceptual Development*, *1*. doi: https://doi.org/10.1007/978-1-4684-2619-9_5

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

Van Hedger, S. C., & Nusbaum, H. C. (2018). Individual differences in absolute pitch performance: Contributions of working memory, musical expertise, and tonal language background. *Acta Psychologica*, *191*, 251-260.

Yu, K., Chen, Y., Wang, M., Wang, R., & Li, L. (2022). Distinct but integrated processing of lexical tones, vowels and consonants in tonal language speech perception: Evidence from mismatch negativity. *Journal of Neurolinguistics*, *61*.

A Appendix

Table A.1: Table of used experiment stimuli

Dutch		**Pseudowords**	
Same	**Different**	**Same**	**Different**
zit zit (sit)	gun kun (grant can)	gep gep	sut zut
dal dal (valley)	bek dek (beak deck)	sut sut	mel nel
gok gok (guess)	dal pal (valley firm)	mel mel	mef nef
mop mop (joke)	bak dak (bin roof)	dos dos	dos tos
tas tas (bag)	gat kat (hole cat)	ves ves	sef zef
bak bak (bin)	vin win (fin win)	sil sil	vom wom
gun gun (grant)	vel wel (membrane well)	rag rag	vus wus
dek dek (deck)	vis wis (fish delete)	sek sek	sek zek
dom dom (stupid)	das pas (tie pass)	vus vus	muk nuk
dan dan (then)	dam tam (dam tame)	gel gel	ris tis
pas pas (pass)	gok kok (guess cook)	mem mem	gel kel
win win (win)	bom dom (bomb stupid)	mef mef	sil zil
vel vel (membrane)	mop nop (joke stud)	muk muk	mem nem
tam tam (tame)	dan pan (then pan)	sor sor	ves wes
vis vis (fish)	wit zit (white sit)	vom vom	sor zor
Akan		**Tone**	
Same	**Different**	**Same**	**Different**
nyà nyà	kyè kyé	713Hz 713Hz	1893Hz 1704Hz
bɔ́ bɔ́	wèn wén	396Hz 396Hz	1778Hz 1600Hz
dà dà	hyɜ̀ hyɜ́	1089Hz 1089Hz	671Hz 504Hz
tɔ̀n tɔ̀n	bù bú	545Hz 545Hz	1129Hz 1328Hz
bà bà	hù hú	660Hz 660Hz	551Hz 822Hz
hù hù	kɔ̀ kɔ́	352Hz 352Hz	951Hz 637Hz
nómm nómm	tɔ̀n tɔ́n	903Hz 903Hz	577Hz 386Hz
dì dì	nyà nyá	908Hz 908Hz	601Hz 707Hz
hyɜ́ hyɜ́	pàm pám	473Hz 473Hz	1055Hz 1002Hz
wè wè	duà duá	440Hz 440Hz	941Hz 706Hz
kɔ̀ kɔ̀	dà dá	756Hz 756Hz	384Hz 404Hz
tó tó	wià wiá	559Hz 559Hz	342Hz 360Hz
twí twí	twì twí	998Hz 998Hz	595Hz 535Hz
kán kán	tò tó	376Hz 376Hz	1094Hz 1287Hz
wiá wiá	wɔ̀ wɔ́	449Hz 449Hz	611Hz 458Hz

B Appendix

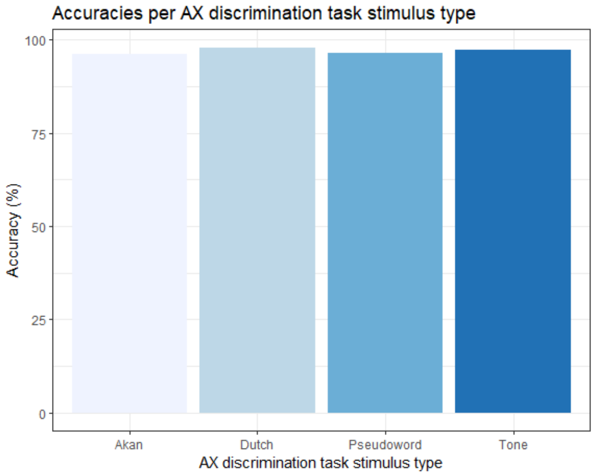

Figure B.1: The accuracies from the AX discrimination task per stimulus type.

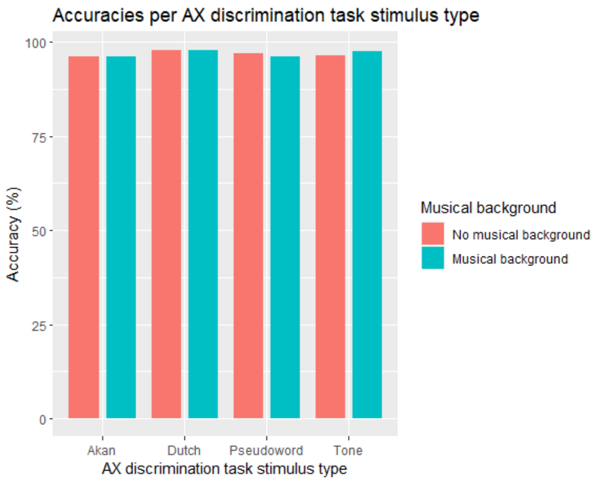

Figure B.2: The accuracies from the AX discrimination task per stimulus type, taking into account musical background.

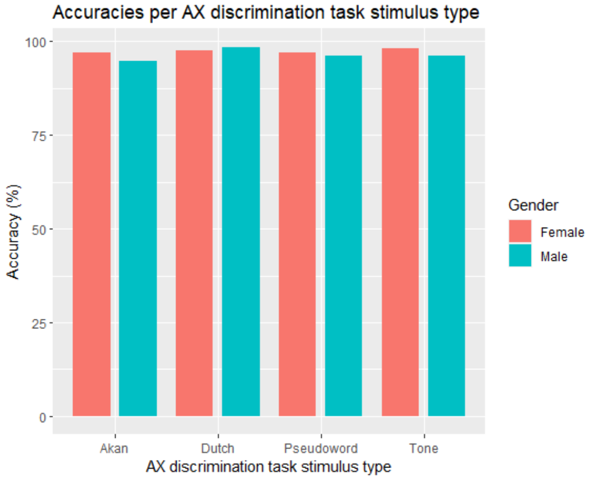

Figure B.3: The accuracies from the AX discrimination task per stimulus type, taking into account gender.