Non-native vowel perception of Swedish and Japanese native speakers

MARIA-CHRISTINA MICALLEF

KURITA MASAKI

Date: March 20, 2023

Supervisor: Axel Ekström
Examiner: Jens Edlund

Abstract

Individuals are better able to recognize that two sounds are distinct if they are separate phonemes in their native language, i.e. listeners exhibit categorical perception, being better able to distinguish sounds that fall in different categories. The phoneme inventory of one's native language thus influences phoneme perception. The project will investigate the effect of native language vowel inventory on vowel perception, with a focus on where boundaries between vowels fall in languages with a small and those with a large vowel inventory. Swedish was chosen as a language with many vowels (9 phonemes, consisting of a long and short allophone of different vowel quality, and Japanese as an example of a language with few vowels (5 phonemes whose long and short versions have the same vowel quality).

Contents

1 Introduction **1**

2 Background **3**
2.1 The role of formants in vowel perception 3
2.2 Frameworks on vowel perception 4

3 Method or Methods **7**

4 Code **15**

5 Results, Analysis and Discussion **17**

References **21**

A Something Extra **23**

List of Figures

2.1 Caption 4

5.1 Graph of the number of times each artificial vowel-phoneme pair were considered to be the same vowel. Key: Swedish /a/: sa, Japanese /a/: ja, //: A, Swedish /o/: so, Japanese /o/: jo, /u/: u. The two vowels in each pair are separated by spaces. If the two formants of a vowel differ, they are separated by a +. If a formant is at the midpoint between the formants of two phonemes, it is represented by the formants of the two phonemes separated by a dot. E.g. u+u.so u compared a vowel with the F1 of an /u/ and the F2 at the midpoint between an /u/ and a Swedish /o/, with the vowel /u/. Each section delineated by black lines contains a single Swedish or Japanese phoneme x being compared with a number of artificial vowels whose formants are between phoneme x and another phoneme e.g. the first section compares vowels whose formants vary between /u/ and the Swedish /o/ being compared with /u/. 18

List of Tables

5.1 Distinct vowel pairs labelled as identical, as % of total responses per language 17

Chapter 1

Introduction

According to Best's Perceptual Assimilation Model, non-native vowels are understood by assimilation into one's native vowel inventory. This project aims to investigate the boundaries between phonemes and how assimilation of a number of back phonemes occurs for a language with a large vowel inventory (Swedish) and one with a small vowel inventory (Japanese), in order to gain a better understanding of phoneme boundaries in languages with a small and with a large phoneme inventory. This will be done by presenting a number of Japanese and Swedish participants with synthesized vowel pairs consisting of the phonemes and vowels created by varying the phonemes' formants so that they lie in-between adjacent phonemes.

Chapter 2

Background

2.1 The role of formants in vowel perception

[1] presented listeners with a series of vowels (iɪɛɑʊuʌɜae) they were asked to identify. The vowels were taken from recordings of actual words but presented in isolation. Some vowels were misclassified far more often than others: 143 out 152 instances of /i/ were correctly classified as such by all listeners, while only 9 of 152 instances of a ɑ were. They note that misclassified vowels were normally labelled as vowels which were adjacent to them in terms of first and second formants, being closest in the loop in the following figure they presented:

The vowels /i/ /ɜ/, /æ/ and /u/ were the most intelligible to listeners. Following American dialectical trends where vowels become more close, listeners tended to misclassify vowels as more open than they were, seemingly 'correcting' for accent. Thus, native speakers of a language are not equally good at distinguishing all the phonemes in that language. It is unclear from this paper whether these specific differences were specific to US native English speakers (which made up the vast majority of listeners) or could be generalised across languages. Secondly, first and second formants play a major role in phoneme recognition.

Swedish consists of 18 vowels, but these are divided into allophone pairs consisting of long and short vowels. Swedish /u/ /o/ and /ɑ/ are long vowels, but /a/ is the short form of /ɑ/ [2]. Japanese has no such qualitative distinctions between its long and short vowels [3].

[4] examined the role of formants and duration in differentiating between 3 long-short Swedish vowel pairs, /i/ and /ɪ, /o/ and /ɔ/, and /ɑ/ and /a/. They did so by, for each vowel pair, synthesizing vowels that varied in duration

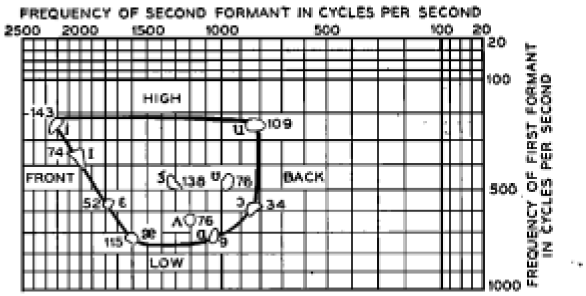

Fig. 3. Vowel loop with numbers of sounds unanimously classified by listeners; each sound was presented 152 times.

Figure 2.1: Caption

between that of the long and short variant, and also varied in terms of F1 and F2. Participants were played a synthesized word containing a vowel and chose which of two real words, one containing a long vowel and the other a short vowel (e.g. vit and vitt), had the same vowel as the synthesized word. They found that duration plays the largest role in differentiating between long and short vowels, particularly when differentiating between /i/ and /ɪ/ and between /o/ and /ɔ, but that the first two formants also play a role in identifiying /a/. As such, since the role of duration is still dominant in differentiating /ɑ from a, this may make it more difficult for Swedish listeners to distinguish between /a/ and /ɑ. However, a difference in formants does exist.

2.2 Frameworks on vowel perception

[5] examined the perception of non-native vowels in relation to Best's Perceptual Assimilation Model (PAM) and Natural Referent Vowel (NRV) frameworks. NRV states that it is easier to detect a vowel change from a less to a more peripheral vowel compared to the other way round. NRV is not particularly useful for this project, as the vowels being examined are all peripheral, being extremely back or extremely open. However, PAM is. According to PAM, people tune themselves to recognizing differences that signal different phonological categories and ignoring differences that do not do so. Thus, people's ability to distinguish between non-native vowels depends on whether they are perceived as different phonological categories. They are most able to distinguish between sounds that are perceived in this manner, which undergo two-category assimilation (TC), are less able to distinguish between

sounds that are assimilated as a good and poor version of the same category (category-goodness assimilation (CG)) and are least able to distinguish sounds assimilated as equally good or equally poor versions of the same category. It should be easy to differentiate between sounds where one is categorized as a particular phonological category and the other is unable to be categorized as belonging to any phonological category within the language (uncategorized-categorized assimilation). If both sounds are uncategorized (uncategorized-uncategorized assimilation), the discrimination varies from poor to very good. [6] examined English speakers' perception of German vowels and found that participants' ability to discriminate between vowels followed Best's model. Participants were able to distinguish between non-native vowels well, but not as well as native speakers.

Chapter 3

Method or Methods

The experiment is a quantitative pilot study. The long Swedish vowels (/ɑ/ /o/ and /u/) were taken from [7], who recorded 26 male speakers saying Swedish vowels in isolation and found the mean formants. The short Swedish central /a/ was based on the results of [8], who measured the formants of short Swedish vowels. They recorded four Swedish male speakers reading words containing short vowels (including /a/) twice and four mean values for each vowel: for those found in the first syllable of a two-syllable word where the syllable is stressed, those in the same context but with the syllable unstressed, those found in the second syllable with the syllable stressed, and those found in the second syllable with it unstressed. As we were unable to find research that provided formants of short vowels recorded in isolation, we calculated the mean of the two average vowels with the syllable stressed.

Calculation of mean Swedish /a/		
stressed vowel	F1	F2
initial position	653	1162
final position	662	1186
average	657.5	1174

The Japanese vowels /o/ and /a/ were taken from a Japanese publication * , which recorded 3 male speakers. The average formants of the two vowels were found.

* Japanese text does not appear in references. As such, please find the paper cited below: 今野英明, 外山淳, 新保勝, and 村田和美. "無声母音のホルマント周波数と音韻性に関する検討." 日本音響学会誌 50, no. 8 (1994): 623-630.

Caclulation of mean Japanese /o/ and /u/				
	/o/		/a/	
Speaker	F1	F2	F1	F2
Speaker A	515	798	701	1206
Speaker B	538	854	826	1374
Speaker C	548	911	782	1306
Averages (to 3 decimal points)	533.667	854.333	769.667	1295.333

Formants used to synthesize vowels		
Vowel	F1	F2
/u/ (Swedish)	290	595
/o/ (Swedish)	390	690
/o/ (Japanese)	533.667	854.333
/ɑ/ (Swedish)	600	925
/a/ (Swedish)	657.5	1174
/a/ (Japanese)	769.667	1295.333

Due to the major differences in formants between the Swedish and Japanese variants of what are ostensibly the same vowels (/o/ and /a/), they are treated as separate.

The 6 vowels form a continuum, with both formants gradually increasing. Artificial vowels were synthesized in the spaces between each pair of adjacent vowels, by varying the f1 and f2, separately and together, to the midpoint between each adjacent vowel pair. [9] notes that people do not perceive pitch linearly. Thus, calculating differences in pitch in terms of Hertz is illogical. The paper examined various psycho-acoustic scales for measuring changes in frequency, where each increase by x on the scale corresponds with the same change in pitch perceived by humans. Of those examined (semitones, mels, Bark and ERB-rate) showed semitones to be the best measure of human pitch perception. As such, the midpoint between vowel formants was calculated in terms of semitones, using this website (https://kassaflutes.com/articles/calculation-of-semitone-intervals). The two formants were inputted into the page, which returned the distance between them in semitones. This distance was divided by 2. The lower formant and the halved distance were then inputted, and the page returned a frequency in Hertz which was halfway between the two formants in terms of semitones. For each adjacent vowel pair (Swedish /u/ and Swedish /o/, Swedish /o/ and Japanese /o/, Japanese /o/ and Swedish /ɑ/, Swedish /ɑ/ and Swedish /a/, and Swedish /a/ and Japanese /a/), pair were created with the first and second formant either the same as that of the lower frequency vowel (L), the same as that of the higher frequency vowel (H) or at the midpoint (m), for 5 vowels per adjacent vowel pair, as follows:

Artificial vowels generated per adjacent vowel pair		
	F1	F2
1	L	M
2	M	L
3	M	M
4	M	H
5	H	M

In addition, the five vowel phonemes were also synthetically generated, resulting in a total of 31 vowels, found below. Vowel names consists of the vowel the first formant corresponds with, the + sign, and the vowel the second formant corresponds with. Formants at the midpoint between two vowels are represented by a vowel, a point (.) and the second vowel. A letter s or j indicates the Swedish or Japanese vowel respectively. Thus, u+u.so refers to a vowel whose first formant is the same as that of /u/ and the second the same as the midpoint between the second formant of /u/ and that of the Swedish /o/.

Generated vowels		
Vowels	f1	f2
u	290	595
u+u.so	290	640.733
u.so+u	336.313	595
u.so+u.so	336.313	640.733
u.so+so	336.313	690
so+u.so	390	640.733
so	390	690
so+so.jo	390	767.773
so.jo+so	456.218	690
so.jo+so.jo	456.218	767.773
so.jo+jo	456.218	854.333
jo+so.jo	533.667	767.773
jo	533.667	854.333
jo+jo.ɑ	533.667	888.968
jo.ɑ+jo	565.858	854.333
jo.ɑ+jo.ɑ	565.858	888.968
jo.ɑ+ɑ	565.858	925
ɑ+jo.ɑ	600	888.968
ɑ	600	925
ɑ+ɑ.sa	600	1042.123
ɑ.sa+ɑ	628.086	925
ɑ.sa+ɑ.sa	628.086	1042.123
ɑ.sa+sa	628.086	1174
sa+ɑ.sa	657.5	1042.123
sa	657.5	1174
sa+sa.ja	657.5	123.222
sa.ja+sa	711.398	1174
sa.ja+sa.ja	711.398	123.222
sa.ja+ja	711.398	1295.333
ja+sa.ja	769.667	123.222
ja	769.667	1295.333

Vowel pairs were then generated. Each phoneme was compared to all the artificial vowels adjacent to it, creating 50 pairs. Adjacent phonemes were also compared to each other, generating an additional 5 pairs. An identical pair of each synthesized vowel was generated in order to mix different vowel pairs with ones that were identical in the experiment, for an additional 31. Finally, 5 very distinct vowel pairs were created in order to confirm that participants' sound system was working well enough for their choices to be meaningful, i.e. that the results obtained are valid. Thus, 91 pairs total were generated.

Vowel pairs used in experiment		
Pairs	vowel 1	vowel 2
Very distinct pairs, to check for faulty audio	jo	ja
	u	sa
	so	ja
	u	ja
	so	sa
artificially generated and naturally occurring synthesized pairs	u+u.so	u
	u.so+u	u
	u.so+u.so	u
	u.so+so	u
	so+u.so	u
	u+u.so	so
	u.so+u	so
	u.so+u.so	so
	u.so+so	so
	so+u.so	so
	so+so.jo	so
	so.jo+so	so
	so.jo+so.jo	so
	so.jo+jo	so
	jo+so.jo	so
	so+so.jo	jo
	so.jo+so	jo
	so.jo+so.jo	jo
	so.jo+jo	jo
	jo+so.jo	jo
	jo+jo.ɑ	jo
	jo.ɑ+jo	jo
	jo.ɑ+jo.ɑ	jo
	jo.ɑ+ɑ	jo
	ɑ+jo.ɑ	jo
	jo+jo.ɑ	ɑ
	jo.ɑ+jo	ɑ

	jo.ɑ+jo.ɑ	ɑ
	jo.ɑ+ɑ	ɑ
	ɑ+jo.ɑ	ɑ
	ɑ+ɑ.sa	ɑ
	ɑ.sa+ɑ	ɑ
	ɑ.sa+ɑ.sa	ɑ
	ɑ.sa+sa	ɑ
	sa+ɑ.sa	ɑ
	ɑ+ɑ.sa	sa
	ɑ.sa+ɑ	sa
	ɑ.sa+ɑ.sa	sa
	ɑ.sa+sa	sa
	sa+ɑ.sa	sa
	sa+sa.ja	sa
	sa.ja+sa	sa
	sa.ja+sa.ja	sa
	sa.ja+ja	sa
	ja+sa.ja	sa
	sa+sa.ja	ja
	sa.ja+sa	ja
	sa.ja+sa.ja	ja
	sa.ja+ja	ja
	ja+sa.ja	ja
adjacent natural vowels	u	so
	so	jo
	jo	ɑ
	ɑ	sa
	sa	ja
identical pairs	u	u
	u+u.so	u+u.so
	u.so+u	u.so+u
	u.so+u.so	u.so+u.so
	u.so+so	u.so+so
	so+u.so	so+u.so
	so	so
	so+so.jo	so+so.jo
	so.jo+so	so.jo+so

	so.jo+so.jo	so.jo+so.jo
	so.jo+jo	so.jo+jo
	jo+so.jo	jo+so.jo
	jo	jo
	jo+jo.ɑ	jo+jo.ɑ
	jo.ɑ+jo	jo.ɑ+jo
	jo.ɑ+jo.ɑ	jo.ɑ+jo.ɑ
	jo.ɑ+ɑ	jo.ɑ+ɑ
	ɑ+jo.ɑ	ɑ+jo.ɑ
	ɑ	ɑ
	ɑ+ɑ.sa	ɑ+ɑ.sa
	ɑ.sa+ɑ	ɑ.sa+ɑ
	ɑ.sa+ɑ.sa	ɑ.sa+ɑ.sa
	ɑ.sa+sa	ɑ.sa+sa
	sa+ɑ.sa	sa+ɑ.sa
	sa	sa
	sa+sa.ja	sa+sa.ja
	sa.ja+sa	sa.ja+sa
	sa.ja+sa.ja	sa.ja+sa.ja
	sa.ja+ja	sa.ja+ja
	ja+sa.ja	ja+sa.ja
	ja	ja

117.9 was used as the F0 for the synthesized vowels, in reference to [10], who recorded 20 male and 20 female speakers and found the mean male F0 to be 117.9 Hz.

The vowels were synthesized using R. The audio, phonTools, and tuneR libraries were used to generate synthesized audio files of two vowels separated by noise.

The target population consisted of Swedish and Japanese native speakers. Convenience sampling was used due to time constraints and this project's status as a student project. The sample was made up of 5 Japanese native speakers and 5 Swedish native speakers. The study's reliability increases with the number of participants. As such, a followup study with more time should have more participants to increase reliability.

The experiment made in PsychoPy was sent to participants to answer on their PC. We determined that there was no risk of physical or psychological to participants. Participants were offered 20SEK for their time. At first names were asked for as a control measure to make sure the people we asked by convenience sampling were answering the questionnaire. However, this request was then removed from the experiment because we became aware that it impacts anonymity. In a large scale experiment where keeping track of responses were more important, a unique code could be assigned to each respondent to keep track of responses while respecting participants' privacy.

The experiment to investigate vowel perception was created in PsychoPy.It consists of 92 routines which play a different audio file respectively. The first routine is used as a trial to familiarize participants with the answering method. Each routine plays an audio file containing a vowel pair after count down from 3 to 1 displayed on the screen. After listening to the audio file, the participants are asked to press 'y' on their keyboards if they think they have heard a pair of the same vowels, otherwise they are asked to press 'n'. The experiment takes about 20 minutes.

Chapter 4

Code

The code below is a program to synthesize vowels written in R.

```
library(audio)
library(phonTools)
library(tuneR)

noise1 = rnorm(5000, 0, 0.05) * 0.2

vowel1 = vowelsynth(ffs = c(769.667, 1233.222), fbw = 0.06,
dur = 500, f0 = c(117.9),fs = 10000, verify = FALSE,
returnsound = FALSE, noise1 = .010, noise2 = .030, power = NULL)

vowel2 = vowelsynth(ffs = c(657.5, 1174), fbw = 0.06,
dur = 500, f0 = c(117.9),fs = 10000, verify = FALSE,
returnsound = FALSE, noise1 = .010, noise2 = .030, power = NULL)

vowels = c(noise1, vowel1, noise1, vowel2, noise1)


writesound (vowels, filename = "vowelpairs.wav", fs = 10000)
```

Another code below is part of a program to let users listen to an audio file and ask them to answer whether they have heard a pair of the vowels or different vowels.It is written in Python using PsychoPy.

```
sound_vowels = sound.Sound('mainpairs/joja.wav', secs=-1,
stereo=True, hamming=True,name='sound_vowels')
sound_vowels.setVolume(1.0)
text = visual.TextStim(win=win, name='text',
```

```
    text='Listen␣to␣a␣pair␣of␣vowels.␣Please␣answer
␣␣␣␣whether␣they␣are␣the␣same␣vowels␣or␣not.␣It␣will␣be
␣␣␣␣played␣in␣3␣seconds.',
    font='Open␣Sans',
    pos=(0, 0), height=0.05, wrapWidth=None, ori=0.0,
    color='white', colorSpace='rgb', opacity=None,
    languageStyle='LTR',
    depth=-1.0);
text_2 = visual.TextStim(win=win, name='text_2',
    text='If␣you␣think␣the␣voewls␣were␣the␣same.␣press␣"y"
␣␣␣␣on␣your␣keyboard␣now.␣Otherwise,␣press␣"n".',
    font='Open␣Sans',
    pos=(0, 0), height=0.05, wrapWidth=None, ori=0.0,
    color='white', colorSpace='rgb', opacity=None,
    languageStyle='LTR',
    depth=-2.0);
text_3 = visual.TextStim(win=win, name='text_3',
    text='3',
    font='Open␣Sans',
    pos=(0, 0), height=0.05, wrapWidth=None, ori=0.0,
    color='white', colorSpace='rgb', opacity=None,
    languageStyle='LTR',
    depth=-3.0);
text_4 = visual.TextStim(win=win, name='text_4',
    text='2',
    font='Open␣Sans',
    pos=(0, 0), height=0.05, wrapWidth=None, ori=0.0,
    color='white', colorSpace='rgb', opacity=None,
    languageStyle='LTR',
    depth=-4.0);
text_5 = visual.TextStim(win=win, name='text_5',
   text ='1'
   font ='Open␣Sans'
   pos =( 00 ) height = 005 wrapWidth = None ori = 00
   color ='white' colorSpace ='rgb' opacity = None
   languageStyle ='LTR'
   depth =-50 );
key_resp = keyboard.Keyboard()
```

Chapter 5

Results, Analysis and Discussion

Swedish speakers were far more likely to classify different vowels as being the same, while Japanese speakers were far more likely to classify the same vowel as being different. Thus, cultural differences likely affected results.

Due to the small number of participants, scores were grouped. X<->y refers to all artificial vowels whose formants ae the same as or between vowels x and y. Scores refer to the number of times aritficial vowels lying in this span were labelled as being the same vowel as the phoneme presented with them. Only scores that fall below 33.33% were taken into consideration, meaning that vowels in that span were generally labelled by participants as different from the adjacent vowel, and those falling above 66.66%, meaning the vowels were generally considered the same as the adjacent vowel. Both Swedish and Japanese speakers considered the Swedish and Japanese /o/ to be distinct, suggesting that the two vowels are not assimilated into the same category for either group, or at the very least that they are not considered a good example of their respective /o/ vowel. Both Swedish and Japanese participants considered vowels lying between Japanese /o/ and Swedish /ɑ/ to be similar to Swedish /ɑ/. Japanese participants generally considered the formant space between Japanese /o/ and Swedish /ɑ/ to fall under the category of an /ɑ/. Swedish participants considered the span between /ɑ/ and Japanese /o/ to fall under /ɑ/ and the span between it and Swedish /a/ to NOT fall

	Swedish	Japanese
Different vowels classified as the same	44	34
The same vowel classified as different	3.870967742	9.032258065

Table 5.1: Distinct vowel pairs labelled as identical, as % of total responses per language

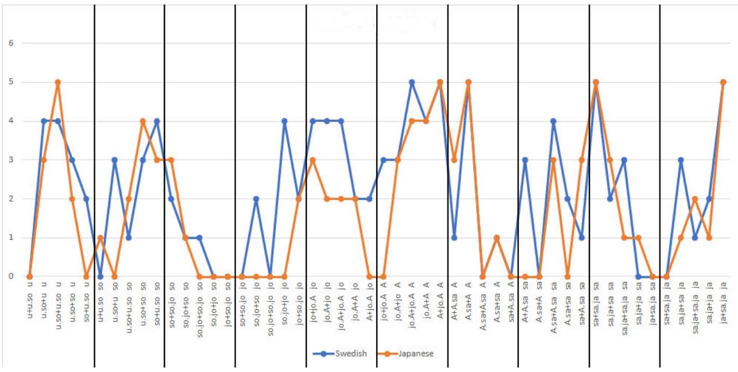

Figure 5.1: Graph of the number of times each artificial vowel-phoneme pair were considered to be the same vowel. Key: Swedish /a/: sa, Japanese /a/: ja, //: A, Swedish /o/: so, Japanese /o/: jo, /u/: u. The two vowels in each pair are separated by spaces. If the two formants of a vowel differ, they are separated by a +. If a formant is at the midpoint between the formants of two phonemes, it is represented by the formants of the two phonemes separated by a dot. E.g. u+u.so u compared a vowel with the F1 of an /u/ and the F2 at the midpoint between an /u/ and a Swedish /o/, with the vowel /u/. Each section delineated by black lines contains a single Swedish or Japanese phoneme x being compared with a number of artificial vowels whose formants are between phoneme x and another phoneme e.g. the first section compares vowels whose formants vary between /u/ and the Swedish /o/ being compared with /u/.

under /ɑ/. This can be understood in terms of PAM: as both /ɑ/ and /a/ are phonemes in Swedish, speakers are sensitive to distinctions between them. On the other hand, Japanese // is not a phoneme in Swedish, so that formant space is more easily assimilated into a /ɑ/ by Swedish speakers. This does not explain why Swedish speakers had difficulty being consistent in their categorisation of vowel pairs in the span between Swedish /o/ and /u/, despite both being phonemes in Swedish.

Vowel pairs, grouped	Swedish	% labelled identical	Difference in % identical labels between 2 adjacent vowels	Japanese	% labelled identical	Difference in % identical labels between 2 adjacent vowels
u<->so u	13	52		10	40	
u<->so so	11	44	8	10	40	0
so<->jo so	4	16		4	16	
so<->jo jo	8	32	-16	2	8	8
jo<->A jo	16	64		9	36	
jo<->A A	20	80	-16	16	64	-28
A<->sa A	7	28		9	36	
A<->sa sa	10	40	-12	6	24	12
sa<->ja sa	10	40		10	40	
sa<->ja ja	11	44	-4	9	36	4

	Swedish participants				Japanese participants			
	Like both adjacent vowels	Like v1	like v2		Like both	like v1	like v2	like neither
u<->so	7	4	10	2	6	7	5	20
so<->jo	0	4	8	0	2	4	19	25
jo<->A	13	7	3	4	12	5	4	25
A<->sa	2	5	9	1	8	10	6	25
sa<->ja	1	10	9	2	7	8	8	25

The study was limited by the small number of participants. In the future, the study could be replicated with more participants. A McNemar test could be used to see if the data of the Swedish and Japanese participants is consistent with each other or if there are differences.

References

[1] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *The Journal of the acoustical society of America*, vol. 24, no. 2, pp. 175–184, 1952.

[2] T. Riad, *The phonology of Swedish*. Oxford University Press, 2014.

[3] L. Labrune, *The phonology of Japanese*. Oxford University Press, 2012.

[4] D. Behne, P. Czigler, and K. P. Sullivan, "Swedish quantity and quality: a traditional issue revisited," *Reports fron the Department of Phonetics, Umeå University*, vol. 4, pp. 81–83, 1997.

[5] M. D. Tyler, C. T. Best, A. Faber, and A. G. Levitt, "Perceptual assimilation and discrimination of non-native vowel contrasts," *Phonetica*, vol. 71, no. 1, pp. 4–21, 2014.

[6] L. Polka, "Linguistic influences in adult perception of non-native vowel contrasts," *The Journal of the Acoustical Society of America*, vol. 97, no. 2, pp. 1286–1296, 1995.

[7] G. Fant, G. Henningsson, and U. Stålhammar, "Formant frequencies of swedish vowels," *STL-QPSR*, vol. 4, no. 1969, pp. 26–31, 1969.

[8] L. Nord, "Acoustic studies of vowel reduction in swedish," *Quarterly Progress and status report*, vol. 4, pp. 19–36, 1986.

[9] F. Nolan, "Intonational equivalence: an experimental evaluation of pitch scales," in *Proceedings of the 15th international congress of phonetic sciences*, vol. 39. Barcelona, Spain, 2003, pp. 771–774.

[10] D. Kovačić and E. Balaban, "Voice gender perception by cochlear implantees," *The Journal of the Acoustical Society of America*, vol. 126, no. 2, pp. 762–775, 2009.

Appendix A

Something Extra

For DIVA


```
{
"Author1": {
        "Last name": "Micallef",
        "First name": "Maria-Christina",
        "E-mail": "maria-christina.micallef.8671@student.uu.se",
        "organisation": {"L1": " ",
                       }
        },
"Author2": {
        "Last name": "Masaki",
        "First name": "Kurita",
        "E-mail": "masaki.kurita.5792@student.uu.se",
        "organisation": {"L1": " ",
                       }
        },
"Degree": {"Educational program": ""},
"Title": {
        "Main title": "Non-native vowel perception of Swedish and Japanese native speakers",
        "Language": "eng" },
"Alternative title": {
        "Main title": "",
        "Language": "swe"
},
"Supervisor1": {
             "Last name": "Ekström",
             "First name": "Axel",
             "Local User Id": "axeleks",
             "E-mail": "axeleks@kth.se",
             "organisation": {"L1": " ",
                            "L2": "Speech, Music and Hearing" }
          },
"Examiner1": {
             "Last name": "Edlund",
             "First name": "Jens",
             "E-mail": "edlund@speech.kth.se",
             "organisation": {"L1": " ",
                            "L2": "Speech, Music and Hearing" }
          },
"Other information": {
"Year": "2023", "Number of pages": "1,23"}
}
```