FACULTEIT DER GEESTESWETENSCHAPPEN
BA Taalwetenschap

Adding Music to a Vowel Contrast

Can the learning of an unknown vowel contrast benefit from the addition of musical features to a distributional learning task?

Gisela Govaart ▪ 10089004
Scriptie BA Taalwetenschap
Universiteit van Amsterdam
July 11, 2014

Supervisors: prof. dr. Paul Boersma & dr. Silke Hamann

‘Zowel in muziek als in taal gaat het om elementen die volgens bepaalde regels in steeds weer andere reeksen wordt gerangschikt. Die overeenkomst heeft knappe koppen ertoe verleid te onderzoeken of de hersenen die twee soorten informatie misschien op dezelfde manier verwerken. Dat wil je toch niet weten? Ze zijn er gelukkig nog niet uit.'

- Herman van Veen -

Samenvatting

In deze scriptie is onderzocht of het toevoegen van muzikale elementen aan een training met een bimodale distributie van een onbekend klinkercontrast het leereffect van de training vergroot, en dus zorgt voor betere resultaten bij een discriminatietaak. 40 participanten hebben geluisterd naar een bimodale distributie van /ɞ/ and /ɜ/, in twee verschillende groepen. Eén groep luisterde naar de klinkers in een spraakconditie, waarin de klinkers aangeboden werden op intonatiecontouren; de andere groep luisterde naar de klinkers in een muziekconditie, waarin de klinkers aangeboden werden op melodieën. Vervolgens zijn beide groepen getest op hun vaardigheid om de doelklinkers te discrimineren, in een AXB-test. Er was geen significant verschil tussen de gemiddelden van de twee groepen; het kan dus niet bevestigd worden dat het toevoegen van muzikale elementen aan een fonetische training het leereffect van de training vergroot. Het is echter mogelijk dat er geen effect van de training is gevonden vanwege verschillen in de training. Er zou ook een non-effect opgetreden kunnen zijn door de aandachtstaak tijdens de training, die er mogelijk voor heeft gezorgd dat de participanten zo gefocust waren op de teltaak dat de doelklinkers niet goed waargenomen zijn. Daarnaast is de muzikaliteit van de participanten getest, met een ingekorte versie van de *Montreal Battery of Evaluation of Amusia* (MBEA), om de mogelijke interactie tussen de MBEA-score en de conditie te onderzoeken. Deze interactie is niet gevonden. Ook is er geen hoofdeffect van de MBEA-score op de AXB score gevonden. Dit komt mogelijk doordat de ingekorte versie van de MBEA geen correcte weergave is van de muzikaliteit van de participanten, aangezien er een negatieve correlatie was tussen de MBEA-score en de zelf-gerapporteerde muzikaliteit.

Abstract

In this thesis I investigated whether the addition of musical features to a distributional learning task can help the learning of an unknown vowel contrast. 40 participants were presented with a bimodal distribution of /ɞ/ and /ɜ/. They were divided in two groups, one of which listened to the vowels in a speech-like condition, in which the vowels were presented on intonation contours; the other group listened to the vowels in a music-like condition, in which the vowels were presented on melodies. Both groups were then tested on their ability to discriminate /ɞ/ and /ɜ/, through an AXB task. A comparison of the means of the two groups showed no significance difference, meaning that it cannot be confirmed that the addition of musical features facilitates the learning of a non-native vowel contrast. However, it might be the case that no difference was found because of differences in the training in the two conditions. The non-effect could also have been caused by the attention task that was carried out during the training, which might have led the participants to focus so much on the counting task that they failed to notice the target vowels. In addition, the participants' musicality was tested, by means of a shortened version of the *Montreal Battery of Evaluation of Amusia* (MBEA), to investigate whether there would be an interaction of the MBEA score and the condition. This interaction was not found, nor was there a main effect of the MBEA score on the AXB score. This might have been due to the possibility that the MBEA did not correctly measure the participants' musicality, since there was a negative correlation of the MBEA score and the self-reported musicality.

Contents

0. Acknowledgements

I would like to thank my supervisors, Paul Boersma and Silke Hamann, for the challenging supervision, with a great combination of a lot of freedom, detailed feedback, and the encouragement to find things out myself. Also, I thank Kateřina Chládková for help with the vowel synthesis, and Dirk-Jan Vet for technical help in the studio. I thank Karlijn Blommers and Jordi Halfman for making the library the nicest place in town, and Jordi for comments on earlier versions of this thesis. Finally, I wish to thank all participants for coming to the Bungehuis to listen to these strange sounds.

1. Introduction

In this chapter, the hypothesis of this thesis will be introduced and explained. In 1.1 the hypothesis will be presented, and then some definitions will be given (1.2). After that, the similarities between music and languages will be pointed out (1.3), and then the location of music and language in the brain will be described (1.4). Next, previous studies on correlations between language skills and musicality will be discussed (1.5), and lastly the concept of distributional learning will be explained (1.6).

1.1 Hypotheses

As can be found in section 1.4 and 1.5, a lot of research has been done on the correlation between musicality and language ability, since music and language seem to share some important characteristics (see 1.3). Therefore, this thesis will explore whether the addition of musical features to a learning task can facilitate the learning process of an unknown vowel contrast. I hypothesize that the addition of musical features increases the learning effect of a training phase with a bimodal distribution (see 1.6) of an unknown vowel contrast, and therefore leads to better results in the discrimination of this contrast. I will thus create a training phase with a bimodal distribution which will be presented to the participants in two conditions: one condition with musical features (music condition), and one condition without musical features (speech condition). I will then test the learning effect of the training, which I expect to be better for the music condition. This hypothesis is based on the following two findings.

First, previous research on language and music shows that musicality correlates positively with different sorts of language tasks (see 1.5.1); I therefore assume that there is a positive connection between the two systems. This positive correlation has also been found for phonetic abilities (see 1.5.2). However, these studies focus on the question of whether good musical abilities correlate with better language abilities; these studies thus try to link two *skills*. This thesis on the other hand, instead of linking the two skills of musicality and discrimination ability, explores whether the addition of *musical features* to a learning task can facilitate the learning process of the task. This has been done before in *Melodic Intonation Therapy* (see e.g. van der Meulen et al. 2012 and Norton et al. 2009), where musical elements are used to emphasize musical aspects that already exist in language, like tone and rhythm. However, I found only two studies that add musical features to a *phonetic* learning task (Karimer 1984 and Lebeveda & Kuhl 2010), and both studies provide no conclusive results on whether adding musical features to a learning task improves phonetic discrimination abilities.

Secondly, similarities in the brain areas involved with the systems for music and language have been found (see 1.4). We would expect that by adding musical features to the training, the training phase addresses not only the language areas in the brain, but the music areas as well; it can therefore be hypothesised that the input will be processed more profoundly. Also, Kolinsky et al. (2009) suggest that the music system and the language system might share attentional capacities; therefore the addition of music might help to allocate attention to the accurate brain area.

Additionally, the hypothesis of this thesis can be supported by the OPERA hypothesis (Patel 2011). According to the hypothesis, musical education improves the neural encoding of language, due to five conditions in the neural networks of speech and music: "(1) **O**verlap: there is anatomical overlap in the brain networks that process an acoustic feature used in both music and speech (e.g., waveform periodicity, amplitude envelope), (2) **P**recision: music places higher demands on these shared networks than does speech, in terms of the precision of processing, (3) **E**motion: the musical activities that engage this network elicit strong positive emotion, (4) **R**epetition: the musical activities that engage this network are frequently repeated, and (5) **A**ttention: the musical activities that engage this network are associated with focused attention." (Patel 2011).

In addition to the above hypothesis, it will be investigated whether an interaction of the participants' musicality and the condition can be found, which would indicate that musical participants benefit more from the addition of musical features. As said above, positive correlations between musicality and language ability have been found (see 1.5), and moreover, research has shown that the part of the brain where both language and music are processed is better developed in musicians than in non-musicians (Sarroff 2009, Schlaug et al. 1995, Schlaug et al. 2005). Therefore, I hypothesize that participants who score high on the *Montreal Battery of Evaluation of Amusia* (see 2.2.1.3) will benefit more from the musical features in the music condition, and will therefore obtain better results in the discrimination of the vowel contrast. Finally, it will be investigated whether the number of languages spoken might have an influence on the ability to discriminate the vowel contrast, since it might be the case that people who speak more languages, have more experience in discriminating vowels that are not present in their mother tongue.

1.2 'Musical features' and 'musicality' defined

In this thesis, a distinction will be made between the terms 'musical features', which refers to certain properties of sound, and 'musicality', which is a skill that people can have to different degrees. Musical features are those properties that make us perceive sound as music, e.g. tone, rhythm, volume and pauses. Surely, musicality cannot be defined extensively. In this thesis, however, a person is considered to be musical if (s)he has some ability to distinguish relative tone frequencies, and has some kind of sense for rhythm. Using this definition, an estimation of a participant's musicality can be made by means of the *Montreal Battery of Evaluation of Amusia* (see 2.2.1.2).

1.3 Similarities between language and music

It is not surprising that the connection between music and language has been researched relatively frequently in the last decades, for there are some striking similarities between the two functions. An important similarity has been brought to light by Lerdahl and Jackendoff (1983), and later by Hauser et al. (2002): the phonemes in language and the notes in music are both elements that can be recombined into structures that have greater meaning than the sum of these elements. Moreover, both systems can recombine these elements, in accordance with a (linguistic or musical) grammar, into an infinite number of new sentences or melodies; this is called *duality of patterning*. Both language and music thus consist of small elements that are ordered hierarchically; in music, notes are ordered so that they form a composition, in language, phonemes are ordered to form a discourse[1]. However, as Sundberg (1991: 442) pointed out, there is a difference to be found in the objective of the structures. In language, structure is created to convey meaning, whereas in music, the structure is an objective in itself. This means that, even though both music and prosody in language[2] can convey emotion to listeners, in music this is more saliently present, because language has another main objective, which interferes with the conveyance of emotion: conveying linguistic meaning.

In addition to the structural similarities, more similarities can be pointed out: both systems are related to pitch, volume and stress, tone, rhythm, and pauses (Mora 2000: 147). In both systems these acoustic characteristics are transferred by means of sound systems. Thus, both systems have to do with processing sound; hence, they both use general auditory processing capacities (Repp 1991: 257). In music and language the input has to go through the

---

[1] Assuming that the construction of one sentence already creates a discourse.

[2] Prosody consists of intonation (pitch), tempo (syllable length) and stress (loudness) (Nooteboom 1997).

same auditory path to reach the area where the sounds are processed. So, at least for a while, both systems use the same capacity, which makes them inherently connected. Consequently, speech can be heard as music[3], and the lyrics that music can encompass are analysed as speech. Furthermore, both systems convey messages that are to be decoded by means of one's knowledge of the system; and this knowledge is transferred to children (or adults) through exposure to the system.

Lastly, and importantly, music and language are perceived in a similar manner. Both intervals in music, and phonemes in language are perceived categorically. For example, Western listeners are not sensitive to the interval of a quarter tone, which occurs in exotic music, because this interval is not used in the Western diatonic-chromatic system (Risset 199: 269). This lack of sensitivity has similarities with the way in which we categorize phonemes: if two sounds are not contrastive[4] in their (first) language, listeners are often unable to discriminate the phonemes[5], resulting in a neutralization of the difference. This occurs for example with /r/ and /l/ for Japanese listeners (Aoyama et al. 2004), and for /ɑ/ and /a/ for Spanish listeners (Escudero et al. 2009). The neutralization of the diference is a result of *categorical perception*: people are more likely to hear differences between phoneme categories than within phoneme categories. Also, similarities between infant speech perception and infant music perception can be noted. Infants are born with the ability to discriminate the sounds of all existing languages. Yet, at about 8 months of age, infants have unlearned this ability, because they have specialized to the sound system of their mother tongue. Something similar happens with musical abilities: musically untrained adults are not able to recognize a pitch alternation of four semitones[6] if this alternation causes neither key-violation nor harmonic changes; infants, however, are capable of detecting these changes (Trehub 2003). In an experiment by Schellenberg and Trehub (1999) it was found that adults perform worse in detecting small pitch alternations in unfamiliar scales than in a familiar (major) scale, in contrast with infants, who performed equally well on both familiar and unfamiliar scales[7]. It therefore seems like phonemic contrasts as well as tone contrasts can be

---

[3] See for an example: philomel.com/phantom_words/sometimes.php

[4] Contrastive means that two sounds are able to make a difference in meaning, e.g. the vowels [æ] and [i] in 'hat' and 'heat'.

[5] However, this is not always the case, consider for example the different occurrences of /r/ in Dutch e.g. [ʀ], [r], [ɾ] etc.: these consonants are not contrastive, but Dutch speakers are still able to identify them.

[6] A major third interval.

[7] However, there was a difference in the infants' performance for different kinds of unfamiliar scales, but this difference is not relevant here.

unlearned, provided that they do not convey meaning. Hence, there seems to be a similarity in the way humans learn and unlearn language contrasts and music contrasts.

In sum, language and music seem to share some important characteristics; it can therefore be assumed that similarities and interactions in the processing of music and language can be found. The next section will discuss the similarities in the location of the two systems in the brain; section 1.5 will discuss some previous findings on how musical skills might interact with language skills.

1.4 Language and music in the brain

It has long been known that music and language are both located in the *planum temporale* (PT), a part of the brain that is positioned at the rear part of the auditory cortex, on the superior side of the temporal lobe, as illustrated in figure 1. The PT can be found on both sides of the brain (Sarroff 2009); for most people, the left PT is larger than the right PT, yet it can be the reverse for left-handed people.

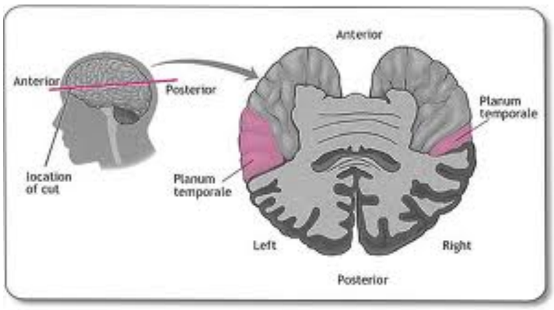

Figure 1. A horizontal section of a normal brain, showing the asymmetry of the planum temporale (OpenLearn LabSpace[8])

Recently more and more studies have focussed on the localization of music and language in the brain, and a more complete view arises. Often, the localization of music in the brain and the localization of language in the brain are investigated separately, yet some studies have investigated the localization of both music and language, in order to look for possible similarities. One of these studies, a PET study by Brown et al. (2006), investigated the location of language and music in the brain, comparing brain activity during the generation of melodies and sentences. It was found that activation was present in nearly

[8] http://www.open.edu/openlearn/education/educational-technology-and-practice/educational-practice/understanding-dyslexia/content-section-4.3.3

identical regions in the brain for language and music [9]. As far as the differences are concerned, they were mostly due to lateralization[10], with a preference for the left hemisphere for language generation. In contrast to previous studies, which found that language is left lateralized, this study found that both music generation and language generation were often bilaterally represented, with significant overlap in brain areas associated with music and brain areas associated with language.

In addition to the research about the differences in *activation* of different brain parts, research focussing on the differences in *cortical structure* in musicians versus non-musicians has been done as well. Different studies (e.g. Sarroff 2009, Chan et al. 1998, Schlaug et al. 1995) suggest that musicians have a better-developed left PT. According to Schlaug et al. (2005), this is not due to pre-existing differences in the brain, which would cause children to decide to learn to play an instrument; this makes it more likely that the differences have to do with engaging in musical activities. In section 1.5 it will be shown that musicians do not only have better developed brain structures in some parts of the brain, but that they actually benefit from this in certain tasks.

In sum, it seems as though music and language use both similar and overlapping brain areas, as well as different areas. Moreover, a certain degree of lateralization has been found, yet neither music nor language can be localized solely in one side of the brain. The attempt of previous studies to localize either music or language in one part or one side of the brain seems too simple for neurological reality. However, it seems sensible to assume that, at least to some extent, music and language use similar areas in the brain. Therefore we might predict that by adding musical features to the training, not only the language areas in the brain are being addressed, but the music areas as well. This could lead to a more profound processing of the signal, resulting in an increased learning effect from the training with musical features.

1.5 Previous studies on correlations between language skills and musicality

It has been suggested that language production and perception can benefit from musicality. A respectable amount of research has been done in this field; this research has focussed mostly on whether musical people show better performance on productive or perceptual language skills. However, this section will also discuss some studies that address whether the addition of musical features to a learning task can aid the learning effect.

---

[9] Brown et al. (2006) names: primary motor cortex, supplementary motor area, Broca's area, anterior insula, primary and secondary auditory cortices, temporal pole, basal ganglia, ventral thalamus, and posterior cerebellum.

[10] The localization of a specific brain function to one side of the brain.

1.5.1 Correlations between musicality and non-phonetic language abilities

Chan et al. (1998) hypothesized that the dissimilarity in cortical structure between musicians and non-musicians springs from a different cortical organisation. Because the left PT is larger for musicians, this would mean that the left PT has better developed cognitive functions for musicians. Because verbal memory is localised in the left PT, Chan et al. tested Chinese students to investigate whether or not musicians have better verbal memory. This appears to be the case: the group that had undergone musical training before the age of 12 showed better verbal memory for spoken words. Piro and Ortiz (2009) found that American children in the age of 7 to 8 had a larger lexicon if they had been taking piano lessons for 3 years. They also found that musically trained children performed significantly better on a semantic task[11]. Hogan and Huesman (2008) also studied verbal memory: they investigated the ability of students to recall strings of words. They found that students who had received musical training for at least 5 years were significantly better in recalling these words. Marin (2009) observed German children of almost 5 years of age (M=4;11, SD=3 months). A part of those children had received musical training, with a mean duration of 4.8 months. These children scored better on the use of morphological rules and the memory of words. They showed increased ability on phonological processing as well, even though these differences were smaller.

Another study, by Jentschke and Koelsch (2009) studied the 'early left anterior negativity' (ELAN) for musically trained children, in order to investigate their syntactical abilities. The ELAN is a neuro-physiological marker measurable with EEG, which is associated with syntactical processing. The ELAN was better developed for musically trained children. Lowe (1998) found that for students who studied French as a second language, oral grammar and text comprehension was significantly better for the group where a music program was integrated in the lessons.

In addition, Schön et al. (2008) performed a study in which it was not the musicality of the participants that was being tested, but the addition of musical features. In their experiment, two groups of 26 French participants heard 21 minutes of nonsense words, presumably part of an unknown language. After the training, the participants were repeatedly presented with two strings, consisting of one 'word' from the training, and one nonword, composed from the syllables in the training; from those two tokens they had to choose the 'word'. If the words were spoken, the participants scored on chance level, whereas they

[11] The children had to choose the right picture for sentence like 'Show me a ball inside a star'.

scored significantly higher than chance level if the words were sung. This might mean that the addition of musical features serves the segmentation of words, especially during the first steps in learning a language.

1.5.2 Correlations between musicality and phonetic and phonological abilities

Regarding the facts that both musicality and phonetic skills are dependent on the capacity to discriminate pitch differences, and that correlations between musicality and other subsystems of language have been found, it may be expected that a correlation can be found between musicality and phonological skills. This section will discuss some research that has found this correlation, as well as two studies that investigated the effect of the addition of musical features to a phonological task.

Pastuszek-Lipińska (2008) studied a group of native speakers of Polish, who were asked to repeat sentences that were presented to them in six different languages they were unfamiliar with. Musicians found the task less difficult than non-musicians, and were able to produce the sentences more fluently. Harrison (1979) found that musically schooled students showed better discrimination and performance of French pronunciation. This corresponds to the findings of Dexter and Omwake (1934), who found a correlation between the ability to detect pitch changes and the pronunciation of French sentences. In addition, Deguchi et al. (2012) found that musicians were better not only in detecting fine pitch changes in tones, but also in speech, in native languages as well as in unknown languages. They suggest that this results from "a more efficient pitch analysis trained by musical experience" (p. 75).

Besides the studies described above, research concerning the effect of the addition of musical features to phonological tasks has been carried out as well. Karimer (1984) tested two groups, which were given four trainings of 20 minutes, spread out over two weeks. The control group listened to minimal pairs in which the target phonemes[12] were highlighted; the experimental group listened to the same sounds in the same context, but here the sounds were presented in songs and rhythmic recitations. The experimental group scored slightly higher[13] on a discrimination test with the target phonemes, and their scores were distributed more evenly around a central score, whereas the scores in the control group were ranged more widely. Likewise, Lebedeva and Kuhl (2010) found that children are better in segmenting syllables if the sentences are sung instead of spoken. They suggest this results from the fact that phonetic information is more easily detectable in combination with clear pitch variation.

[12] Karimer (1984) reports the phonemes as follows: "/sh/, /s/, /z/, /l/, /r/, /c/, /ch/, /p/, /f/".

[13] Though not significantly.

1.5.3 Conclusion

From this section a few conclusions can be drawn. First, there seems to be a positive correlation between musicality and (phonetic) language skills. This leads us to expect that there might be an interaction between the participants' musicality and the condition, meaning that musical participants benefit more from the addition of musical features to the learning task than less musical participants. Second, there have been some studied that investigated whether the addition of musical features to a learning task improves learning effects. These studies are not numerous neither decisive, but seem to indicate that adding musical features to a (phonetic) learning task might facilitate the learning effect of the task.

1.6 Distributional learning: bimodal distributions

Previous research has shown that listeners can be trained on their ability to discriminate an unknown vowel contrast by exposing them to a bimodal distribution of the two vowel categories (Maye & Gerken 2000, 2001); this is called *distributional learning*. Distributional learning is based on the idea of statistical learning: "the ability to track consistent patterns in the input to discover units and structures" (Saffran 2003: 398). A bimodal distribution encompasses a certain number of occurrences of the vowels with different values, along the acoustic continuum (Escudero et al. 2011). On this continuum, two peaks can be found; these are the prototypical realizations of each vowel (see 2.3.1); see figure 2 for a visualization of the bimodal distribution used for this experiment.

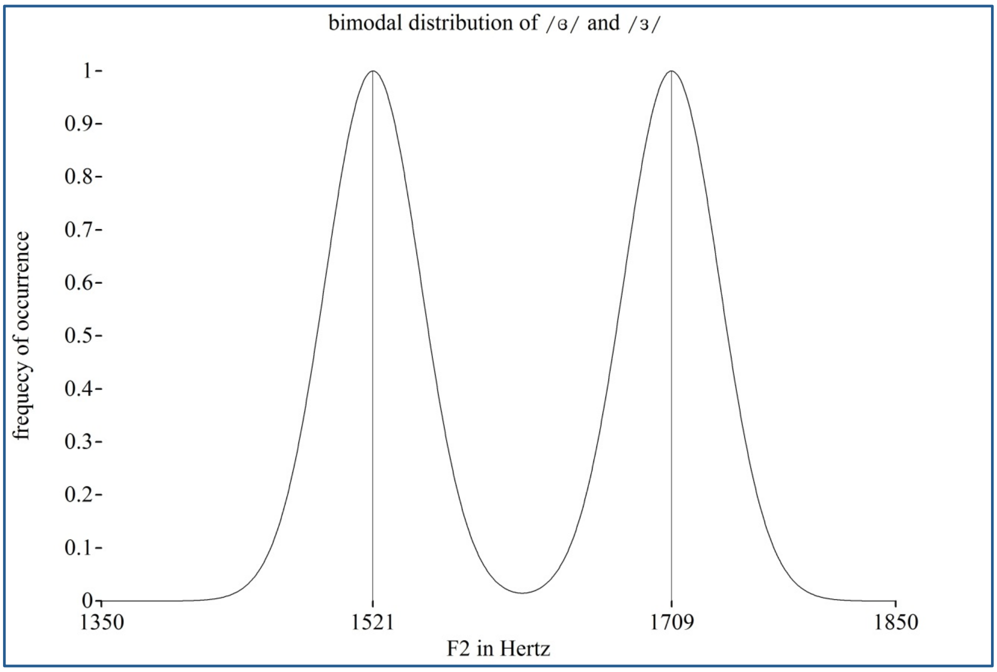

Figure 2. Bimodal distribution of /ɞ/ and /ɜ/ with means of 1521 Hz and 1709 Hz, and a standard deviation of 30 Hz

2. Methods

For this thesis, three experiments have been carried out. A first pilot was performed to find for what second formant (F2)[14] difference participants would be able to discriminate the target vowels correctly 70% of the time, without being exposed to a training phase beforehand. Second, another pilot was carried out, in order to test the experimental design of the final experiment. Ultimately, the final experiment was conducted.

This chapter starts with a description of the participants (2.1), and then it describes the research designs of the final experiment and both pilots (2.2). Section 2.3 will go into the materials, discussing both the sounds as well as the equipment, and section 2.4 describes the procedure.

2.1 Participants

2.1.1 Participants in the pilots

In pilot 1, seven native speakers of Dutch participated. Their ages varied from 19 to 52 years, with a mean age of 26 years. Six subjects had Dutch as their only first language; one subject was bilingual, with Frisian as her other first language. For pilot 2, 14 speakers of Dutch participated; none of these participants had participated in pilot 1. Their age ranged from 21 to 31 years, with a mean age of 23 years. Thirteen subjects had Dutch as their only first language; one subject was bilingual, Frisian being her other first language. For an overview of the participants, see appendix 7.4.1 and 7.4.2.

2.1.2 Participants in the final experiment

In the final experiment, 43 participants took part, none of who had participated in either one of the pilots. Their ages ranged from 20 to 33 years, with a mean age of 24 years. Two of them were bilingual, one subject had Danish as a second mother tongue, and another subject had Japanese as a second mother tongue. The two bilinguals were both removed from the analysis, in order to analyse only native speakers of Dutch. One subject was removed from the analysis due to a technical problem. Of the remaining 40 participants, 27 were female and 13 were male. The participants did not receive any money, but were given a chocolate bar to thank them for participating. For an overview of the participants, see appendix 7.4.3.

---

[14] A formant is a resonance of the vocal tract at a certain frequency. Vowels are mainly characterized by three of those resonances: the first, second and third formant.

2.2 Design

2.2.1 Final experiment

The final experiment consisted of a training phase and two tasks: a discrimination task and a musicality task. The training consisted of four minutes of listening to synthetic vowels, both target vowels /ɞ/ and /ɜ/, and fillers /i/, /a/ and /u/. The fillers were inserted to prevent habituation, i.e. the decrease of attention to a familiar stimulus (Bornstien 2010). The training was presented in two different conditions: a speech-like condition and a music-like condition, in which the vowels were presented on a pitch distribution resembling respectively intonation contours or melodies. In order to make sure that the participants paid attention to the sounds in the training, they were asked to count the occurrences of /a/. /a/ was chosen as the vowel to be counted because it is located in the middle of the vowel triangle (see figure 3), hence, focussing on /a/ would not direct the participants' attention to either side of the vowel triangle, which would cause them to focus more on one of the two target vowels. The vowels in the speech condition were shorter than the vowels in the music condition, /a/ occurred 87 times in the speech condition, and 48 times in the music condition.

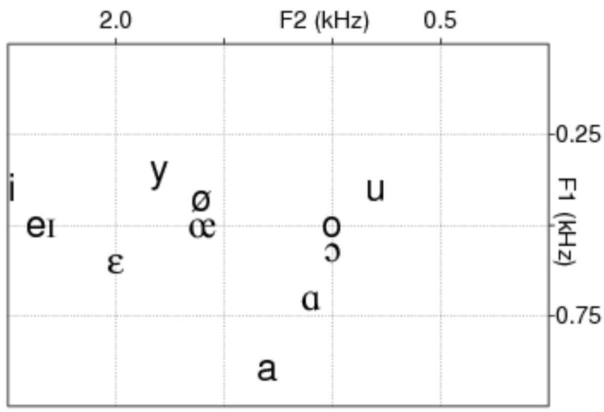

Figure 3. Vowel triangle for Dutch (Welker 2006)

2.2.1.1 The AXB task

After the training phase, a forced choice discrimination task, presented in AXB format, was carried out. For the script used for the AXB-task, see appendix 7.1.5 and 7.1.6. The participants heard three vowels, with an initial silence of 0.8 seconds, an inter-stimulus interval of 0.4 seconds, and a final silence of 0.8 seconds. Then the participants were asked whether they thought the second vowel (X) sounded more like the first (A) or the third vowel

(B). To indicate their choice, they either clicked on the button 'eerste' ('first') or 'derde' ('third'). The participants started with a short practice round, in which they were asked to perform the exercise four times. Since the objective of this practice was just to familiarize the participants with the test design, the vowels /i/, /a/ and /u/ were used, as those vowels can be discriminated easily by native speakers of Dutch.

2.2.1.2 The Montreal Battery of Evaluation of Amusia

The second task the participants were asked to perform was a music task, in which the (a)musicality of the participants was tested. For this task, a shortened version of the *Montreal Battery of Evaluation of Amusia* (MBEA; Peretz et al. 2003) was used. For the scripts used for this task, see appendix 7.1.7. The MBEA consists of six subtests, assessing different music processing skills: scale, contour, interval, rhythm, metric, and music memory. For this thesis, only the skills scale and rhythm were tested. These two skills correspond with the definition of a musical person used in this thesis, namely someone that has some ability to distinguish relative tone frequencies, and has some kind of sense for rhythm. It is therefore plausible that these two subtests will give us a reasonable insight in one's musical abilities, even though the MBEA is designed to test amusicality instead of musicality.

The scale and rhythm subtest of the MBEA consist of 31 exercises per subtest. However, because of the small scope of the present experiment, participants could not be occupied too long. For that reason only 15 exercises per subtest were used for this task; I chose the first 15 exercises, since the melodies were all very similar. Prior to the task, the participants were told that in every exercise they would hear two identical or non-identical melodies. The participants were asked to listen to the two melodies, and decide whether the melodies were the same or different. To indicate their choice, they either clicked on the button 'hetzelfde' ('the same') or 'verschillend' ('different'). The task started with two practice exercises, to familiarize the participants with the task, followed by 15 real exercises, then another two practice exercises, again followed by 15 real exercises.

<u>2.2.2 Pilot 1</u>

Prior to the experiment, two pilots were carried out. The first pilot was conducted to find for what F2 difference untrained participants would be able to discriminate the target vowels correctly 70% of the time; I was thus looking for a /ɵ/-/ɜ/vowel pair with such an F2 difference that the participants could discriminate the vowels 70% of the time. For this purpose, participants heard 45 different vowel pairs with increasing F2 difference (see 2.3.1 for a description of the vowel qualities).

The vowel pairs were tested in an AXB task, similar to the AXB task used in the final experiment. For the script used for this task, see appendix 7.1.4. All 45 pairs were presented four times in AXB triplets, in four different orders, except for the second and third pair. The second pair occurred seven times, and the third pair was not presented at all. Once, a combination of two vowels was presented of which one belonged to the second vowel pair and one to the third vowel pair. This was all due to an error in the script.

The subjects were tested either in the university library or in their own homes. The results were analysed using logistic regression[15], in order to find the threshold of the F2 difference for which participants were able to discriminate 70% of the vowel pairs. Table 1 shows the formula for the logistic regression, as well as the F2 difference of the vowels for which the participants were able to discriminate 70% of the pairs. As can be seen from the table: a difference of 228.14 Hz is given for participant 6KG, while the greatest measured F2 difference in the pilot was 188 Hz. Hence, it can be assumed that this participant did not hear any difference. Therefore, the results of this participant have been removed from the final analysis of the pilot, resulting in a mean difference of 73.08 Hz as the threshold of the F2 difference for which participants were able to discriminate 70% of the vowel pairs.

In figure 4, a scatter plot of the pilot is given. The scatter plot seems to indicate that the F2 difference for which the participants were able to discriminate 70% of the vowel pairs correctly lies somewhere between 75 Hz and 140 Hz; the scatter plot thus suggests that there is more than one F2 difference for which participants are able to discriminate the vowels correctly 70% of the time.

---

[15] "The logistic regression method will find values α, β_{F1} and β_{dur} that optimize $\alpha + \beta_{F1}\,F1_k + \beta_{dur}\,Dur_k = \ln(p_k(/ɛ/)/p_k(/æ/))$, where k runs from 1 to 10, and $p_k(/æ/) + p_k(/ɛ/) = 1$." (Praat manual: Boersma, January 31, 2011).

Table 1. Results from pilot 1

Participant	Formula logistic regression ln (P(correct)/P(incorrect)) ≈ a + b * F2 difference	F2 difference for 70% correct (Hz) F2 difference = (ln (70/30) – a)/b
1JV	≈ -0.700790 + 0.035193 * F2 difference	(0.8473+0.7008)/0.0351 = 44.12
2TJ	≈ -0.706570 + 0.024485 * F2 difference	(0.8473+0.7066)/0.0245 = 63.43
3MJ	≈ -0.602929 + 0.016122 * F2 difference	(0.8473+0.6029)/0.0161 = 90.07
4JJ	≈ -0.831786 + 0.018144 * F2 difference	(0.8473+0.8318)/0.0181 = 92.77
5RP	≈ -0.592587 + 0.015637 * F2 difference	(0.8473+0.5926)/0.0156 = 92.30
6KG	≈ -0.475697 + 0.005808 * F2 difference	(0.8473+0.4759)/0.0058 = 228.14
7RB	≈ -0.251995 + 0.019717 * F2 difference	(0.8473+0.2520)/0.0197 = 55.80
	mean steepness (without 6KG): 0.0219162	mean F2 difference (without 6KG): 73.08 Hz

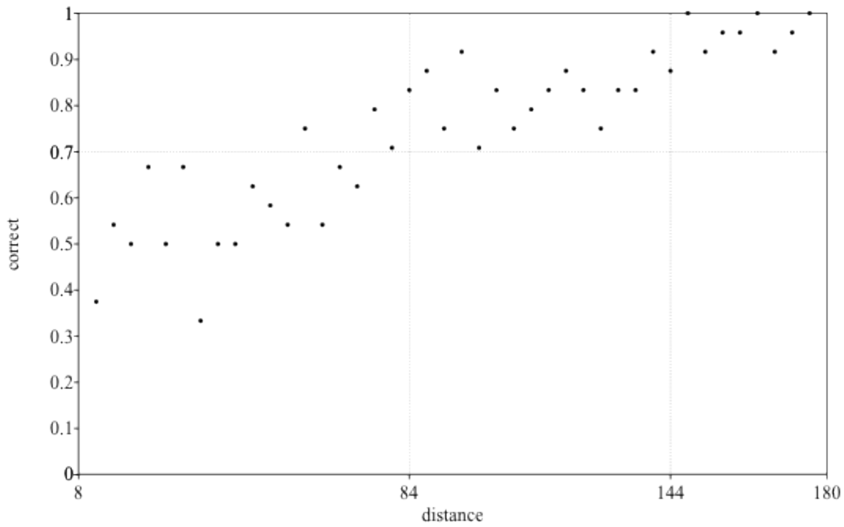

Figure 4. Scatter plot of the results from pilot 1

<u>2.2.3 Pilot 2</u>

In addition to the first pilot, I conducted a second pilot, to test the experimental design of the final experiment. For this pilot, I used the same subtests as in the final experiment. The training and the MBEA contained the same sounds as in the final experiment; only the vowel pairs in the AXB test differed. For the AXB test only one vowel pair was used, namely the one that corresponded to the result of pilot 1 according to the logistic regression, that is, the vowel pair that the participants could discriminate correctly 70% of the time according to the logistic regression (see 2.3.1.2 for a description of the vowel qualities of this pair). However, the scatter plot of the results from pilot 1 suggested that the 70% threshold might not be represented by a single vowel pair, but lies somewhere in between a number of vowel pairs. Pilot 2 was thus conducted to confirm the 70% threshold for the final experiment. Interestingly, the participant performed relatively poorly on the AXB task: on average they discriminated 57.14 % of the vowels pairs correctly, which is only slightly above chance level. Therefore, the vowel pairs for the final experiment were chosen according to the scatter plot analysis of the first pilot.

2.3 Materials

2.3.1 Stimuli

All vowels were synthesised with the Klatt synthesizer in Praat (Boersma & Weenink 2013), using the scripts in appendix 7.1.1 to 7.1.3. The target vowels were synthesized after the vowels /ɘ/ and /ɜ/ from the Slovenian dialect spoken in San Giorgio, a village in Resia, which is situated in the North-East of Italy (Steenwijk 1992). This vowel contrast was chosen because the vowels (almost) only differ in their F2. The mean formant values of the vowels can be found in table 2. The target vowels have been adapted in such way that the first formant (F1) and the third formant (F3) of the two vowels were kept equal, so that the vowels would differ only in their F2. The formant values of the target vowels can be found in table 3.

Table 2. Formant values of the Slovenian vowels (Steenwijk 1992)

	/ɘ/	/ɜ/
F1	566 Hz	549 Hz
F2	1521 Hz	1709 Hz
F3	3209 Hz	3082 Hz

Table 3. Formant values of the target vowels

	/ɘ/	/ɜ/
F1	558 Hz	558 Hz
F2	≤ 1521 Hz	≥ 1709 Hz
F3	3146 Hz	3146 Hz

2.3.1.1 The stimuli for the training

In the training, the target vowels were presented according to a bimodal distribution for F2, with the F2 value of the Slovenian vowels (1521 Hz and 1709 Hz) for the means (that is, the peaks in figure 2) and a standard deviation of 30 Hz. Beside the target vowels, filler vowels were also synthesized (see the scripts in appendix 7.1.2). As fillers, /i/, /a/ and /u/ were used, since these vowels are located at the corners of the vowel triangle, approximately at equal distances from the target vowels (see figure 3). Their formant frequencies were chosen according to Adank et al. (2004: 1732, table I). Since the script that was used to produce the vowels was created to synthesize female vowels, the female frequencies were used. The formants of the filler vowels can be found in table 4. For the synthesis, the F1 of the filler vowels was varied according to a Gaussian distribution with a standard deviation of 15 Hz. This was done in order to prevent participants from listening to token-specific acoustic features that might arise as a result of F0[16]–F1 interactions.

Table 4. Formant values of the filler formants

	/i/	/a/	/u/
F1	294 Hz	912 Hz	286 Hz
F2	2524 Hz	1572 Hz	928 Hz
F3	2911 Hz	2852 Hz	2736 Hz

The vowels were presented either in a speech-like condition or in a music-like condition. For the first condition, intonation contours were used to create a speech-like sound. To create the sounds, I recorded Dutch sentences from ToDI[17] (Gussenhoven et al. 2003). Four of these sentences were analysed in Praat, and their pitch contour was remodelled, using the scripts in appendix 7.1.2. On these four contours, the vowels were inserted randomly, /i/, /a/, and /u/ each occurring 16.6 % of the time, /ɞ/ and /ɜ/ each occurring 25% of the time. These intonation strings had an initial silence of 1 second, an inter-stimulus interval of 0.1 seconds, and a final silence of 0.7 seconds. Subsequently, the four intonation strings were put together 39 times in random order, to form a sound of approximately 4 minutes (242.26 sec). Hereafter, the sounds for condition two were created. For each of the vowels in the four intonation strings in the speech condition, the pitch was matched with the closest tone from the diatonic-chromatic scale (see appendix 7.3 for the tones). Next, four melodies were

[16] Pitch.
[17] Transcription of Dutch Intonation.

created with the tones, using every tone only as often as it occurred in the matched intonation string. On these melodies, the vowels were inserted randomly, /i/, /a/, and /u/ each occurring 16.6 % of the time, /ɘ/ and /ɜ/ each occurring 25% of the time. Each melodic string had an initial silence of 1 second, an inter-stimulus interval of 0.1 seconds, and a final silence of 0.7 seconds. After this, the four strings were put together 40 times in random order, to form a sound of approximately 4 minutes (241.50 sec). For the pitch tracks of the intonation strings and the melodic strings, see figure 5.

The mean pitch of the vowels in the four intonation strings was 254.81 Hz, and the mean pitch of the vowels in the four melodic strings was 254.33 Hz. The mean pitch had to be (almost) the same in the two conditions, because research on thresholds for formant-frequency discrimination has indicated that a high F0 causes a higher threshold than a low pitch, meaning that it is easier to identify a vowel when the F0 is low.

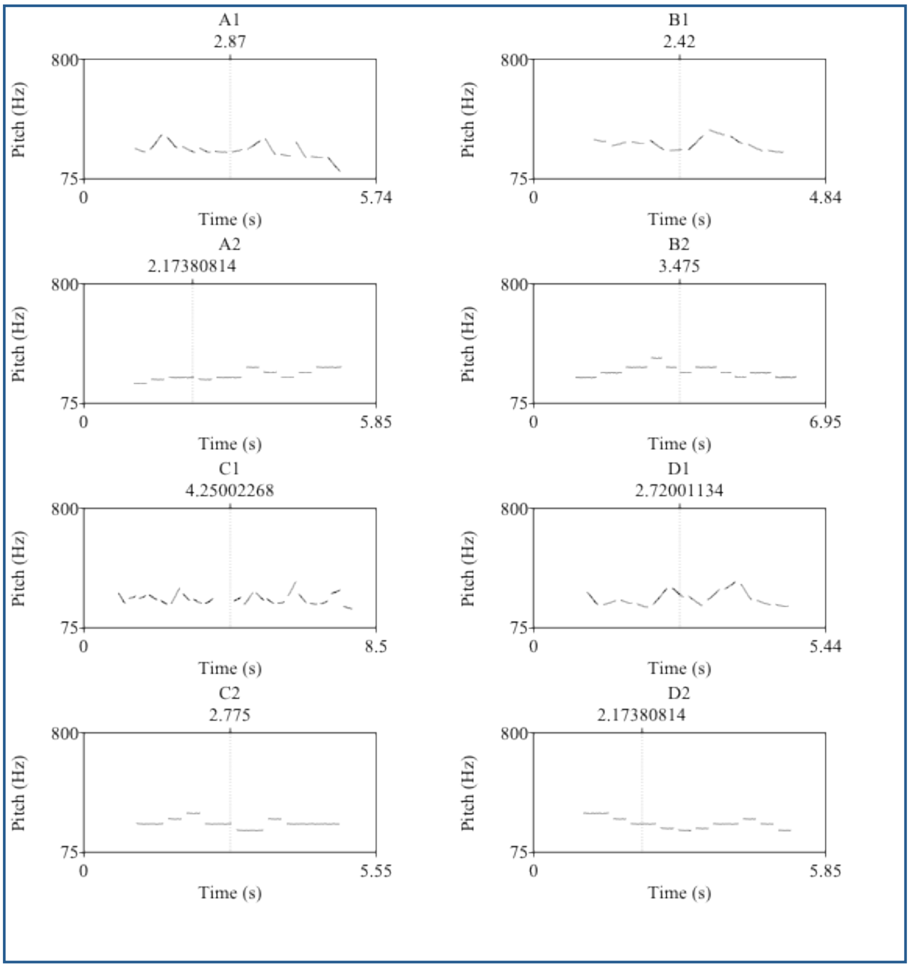

Figure 5. Pitch tracks of the intonation strings (A1, B1, C1, D1) and the melodic strings (A2, B2, C2, D2)

2.3.1.2 The stimuli for the AXB tasks

In pilot 1, participants were presented with 45 different pairs of the target vowels. For /ɜ/, F2 varied from 1709 Hz, which is the value of the Slovenian /ɜ/, to 1621 Hz, going down by steps of 2 Hz. For /ɘ/, F2 varied from 1521 Hz, which is the value of the Slovenian /ɘ/, to 1609 Hz, going up by steps of 2 Hz. Thus, the F2 difference between /ɘ/ and /ɜ/ decreased by 4 Hz with each pair. Consequently, the 45th pair had an F2 difference of 12 Hz, which can be asserted to be small enough to be inaudible, since various studies have reported a difference limen (DL) – just noticeable difference, the smallest difference that is still audible – for F2 far greater than 12 Hz, see for example Mermelstein (1978): "the average DL for F2 is 75 Hz for steady-state vowels" (p. 575). According to this average DL for F2, it also seems sensible to assume that the Slovenian vowel contrast is not inaudible in a discrimination task, since the F2 difference is 188 Hz, which is far greater than 75 Hz. Therefore, no vowel pairs with an F2 difference greater than the F2 difference of the Slovenian vowels have been used in pilot 1. See for an overview of the vowel pairs appendix 7.2.

According to the results of pilot 2 (see 2.2.3), the vowel pairs for the final experiment were chosen according to the scatter plot analysis of pilot 1. Hence, the vowel pairs 12 to 27 (see appendix 7.2), with an F2 difference of 84 Hz to 144 Hz, have been used in the final experiment.

2.3.2 Equipment

Pilot 1 was run in Praat (Boersma & Weenink 2013) on a MacBook Air[18], with Sennheiser earphones[19]. Pilot 2 and the final experiment were also run in Praat, in a soundproof studio. The sounds were presented binaurally on headphones with a sounds pressure level of approximately 55-60 dB for all participants. During the final experiment, the headphones broke, thus the last 17 participants were tested with different headphones; the sound pressure level, however, was kept approximately equal. The AXB task and the MBEA task were presented on a TOBI screen.

2.4 Procedure

For pilot 1, the participant was seated behind the computer, and was told to read the instructions on the screen: "Je krijgt straks steeds drie geluiden te horen. Het is de bedoeling dat je aangeeft waar het tweede geluid het meest op lijkt. Als het tweede geluid het meest lijkt

[18] MBAIR 13.3/1.7/4/128FLASH-NLD.

[19] Sennheiser Headset & Microfoon Mx 470.

op het eerste geluid, klik je op 'eerste', als het tweede geluid het meest lijkt op het derde geluid, klik je op 'derde'. Klik om te beginnen."[20] The exercise was presented 180 times, with a break after every 45 exercises. If the participant stopped after a few exercises, to say that (s)he did not hear a difference, (s)he was told that it was not a problem if (s)he did not hear a difference yet, and that they should just continue listening and concentrate on the task.

In pilot 2 and the final experiment, the participants were first instructed that they were going to listen to four minutes of either speech (intonation contours) or music (melodies), with vowels. They were given a paper and a pen, and were instructed to score the occurrences of /a/ in the sound. After the training, the AXB task was explained, with the same instruction as in pilot 1. Also, they were told that the task could be difficult, and that they should just guess if they were not sure. In pilot 2 the exercise was presented 80 times, with a break after 40 exercises, and in the final experiment the exercise was presented 138 times, with a break after every 50 exercises. Also, in the final experiment a practice round was added. Again, if the participant stopped after a few exercises, (s)he was told that they should just continue and concentrate on the task. This happened with about 30% of the participants. Subsequently, the participants were given instructions for the MBEA task. After the MBEA, the participant was thanked, and informed about the purpose of the experiment.

The participants were assigned to the conditions according to an estimation of their musicality, based on whether they ever sing for a public and/or play an instrument. The participants were classified into three different groups: ‘not musical’, ‘a bit musical’, and ‘musical’. Then, the participants were assigned to either one of the conditions, so that both conditions would contain equally many participants from each group. This was done in order to exclude the possibility that one of the conditions would contain participants with higher scores on the MBEA, since this was expected to influence the results of the AXB task.

[20] Translation: ‘Next, you will hear three sounds a number of times. You should indicate which sound is most similar to the second sound. If the first sound is most similar to the second sound, click ‘first’; if the third sound is most similar to the second sound, click ‘third’. Click to start.’.

3. Results

In this chapter, the results of the final experiment will be presented. A Univariate General Linear Model (GLM) has been applied, with the AXB score as dependent variable, the condition (speech or music) as fixed factor, and the MBEA score as covariate. For the first hypothesis, that the addition of musical features increases the learning effect of a training phase with a bimodal distribution, the GLM enquired whether the difference between the AXB scores for the two conditions is significant. To investigate whether musical participants are better in discriminating the vowels, the GLM enquired whether the MBEA score shows an main effect on AXB score. Then, for the hypothesis that musical people benefit more from the addition of musical features than do less musical participants, the GLM enquired whether there is an interaction between the MBEA score and the condition on the score on the AXB task. Finally, the GLM enquired whether an effect of the number of languages the participants speaks can be found. Additionally, a t-test has been performed in order to rule out that the type of headphone influenced the results on the AXB task.

However, first, the scores on the MBEA for the two conditions were compared, to rule out a pre-existing difference between the two groups, that could have existed, even though the participants were divided over the conditions based on their estimated musicality (see 2.3). The mean AXB scores were 75.91% (between participant SD=14.13) for the speech condition, and 76.40% (between participant SD = 14.15) for the music condition. The data were assumed to be normally distributed for both conditions, since a Kolmogorov-Smirnov test was not significant (p=0.143 for the speech condition, p=0.2 for the music condition). The 95% confidence interval ranges from 69.30% to 82.52% for the speech condition, and from 69.78% to 83.03% for the music condition. The GLM shows no significant difference between the two conditions ($F(1,35)=0.647$, p=0.426).

The mean scores on the MBEA were 82.33% (between participant SD=9.68) for the speech condition, and 87.33% (between participant SD=6.72) for the music condition. On these scores, an independent samples t-test was applied, which showed no significant difference ($t(38)=1.898$, p=0.065). The data of the MBEA task were assumed to be normally distributed for the two conditions, on basis of a Kolmogorov-Smirnov test (p=0.054 for the speech condition, p=0.175 for the music condition). The confidence interval for the means ranges from 69.30% to 82.52% for the speech condition, and from 69.78% to 83.03% for the music condition. The GLM showed no main effect for the MBEA score on the AXB score

(F(1,35)=0.163, p=0.689). Also, the GLM showed no interaction between the scores on the MBEA and the condition on the scores on the AXB task (F(1,35)=0.681, p=0.415).

Taking into account the possibility that the number of languages spoken by the participants could influence their score on the AXB task, the participants were grouped on the basis of the number of languages they reported to speak. They were given 1 point for each language they reported to speak at least 'quite well', and 0.5 points for a language in which they could participate only in basic conversations. This resulted in 7 groups, as can be seen in table 5. The GLM showed no significant main effect from the number of languages spoken on the score on the AXB task (F(1,35)=0.498, and p=0.485).

Table 5. Score on AXB grouped for how many languages the participant speaks

	1.0 *n = 6*	1.5 *n = 1*	2.0 *n = 14*	2.5 *n = 3*	3.0 *n = 9*	3.5 *n = 3*	4.0 *n = 4*
Mean	72.79 %	65.62 %	75.77 %	82.81 %	75.61 %	84.38 %	76.38 %
SD	14.00	-	11.60	14.34	15.72	19.93	19.72

Additionally, a comparison of the means of the groups before and after the switch of headphones has been made. The mean AXB score for the participants with the first headphone (n=23) was 76.44% (between participant SD=12.13); the mean AXB score for the participants with the second headphone (n=17) was 75.78% (between participant SD=16.51). The confidence interval for the means ranges from 71.20% to 81.68% for the group with the first headphone, and from 67.29% to 84.27% for the group with the second headphone. An independent t-test showed no significant difference between the two groups (t(38)=0.145, p=0.886).

4. Discussion

In this chapter, the results of the final experiment will be interpreted (4.1) and discussed (4.2). In addition, some further observations will be discussed (4.3).

4.1 Interpretation of the results

From the results in chapter 3, four conclusions can be drawn:

(1) No significant difference between the speech condition and the music condition has been found, meaning that no effect was found from the addition of musical features to the distributional learning task. Thus, the participants in the music condition did not benefit more from the music-like training than did the participants in the speech condition benefit from the speech-like training.

(2) No interaction between the scores on the MBEA and the condition on the scores on the AXB task was found, meaning that it was not found that musical participants benefit more from the addition of musical features to the distributional learning task than do less musical participants.

(3) No main effect for the MBEA score on the AXB score was found; it was therefore not confirmed that musical participants score better on the AXB task.

(4) No significant main effect from the number of languages spoken on the AXB score was found; meaning that it was not affirmed that participants who speak more languages score higher on the AXB task than do participants who speak fewer languages.

4.2 Possible explanations

<u>4.2.1 No effect from the addition of musical features to a distributional learning task</u>

The most obvious possible explanation for the fact that no results were found in the final experiment, is that adding musical features to a distributional learning task does not aid native speakers of Dutch in learning the non-native vowel contrast /ɘ/ and /ɜ/. This would be an interesting finding, since previous research did find positive correlations for musicality and phonetic skills (see 1.5.2), and so far, no conclusive studies have been carried out to confirm that adding musical features to a learning task can facilitate the acquisition of phonetic skills[21].

[21] Some positive results have been found (see 1.5.2), but not enough to confirm the hypothesis that adding musical features to a learning task actually improves learning results for phonetic skills.

Thus, this finding could possibly indicate that the higher scores on phonetic tasks for musical people can be explained by a difference in cortical structure, since the higher scores do not seem to be due to the activation of the areas in the brain that are related to the processing of music.

However, it was not found that participants with higher MBEA scores also have higher AXB scores, so it could be that there is no effect from musicality on a task like the one carried out in this thesis. Furthermore, a closer look at the studies in 1.5.2 shows that so far, no effect of musicality has been found on a discrimination task exactly like the one in this thesis. However, there is still reason to suggest that a positive correlation between musicality and discrimination skills of non-native vowels can be found. For example, Slevc and Miyake (2006) found a correlation between musical skills and phonological ability (both productive and perceptive) in a second language (L2), suggesting that "musical skills may facilitate the acquisition of L2 sound structure" (p. 675). Even though this study investigated a contrast between two consonants, /r/ and /l/, not two vowels, it can be expected that an effect of musicality on a discrimination task with vowels could be found as well. In addition, production and perception skills are often closely linked, as Flege (2003) points out. Since positive correlations of musicality and production skills have been found, we might assume a positive correlation of musicality and discrimination skills, even though this was not found in this research.

Moreover, there is a possibility that no positive correlation of the MBEA score on the AXB score was found because of the fact that the MBEA scores do not accurately reflect the participants' musicality, since there were some striking results in the MBEA scores. For example, participant 54, who studies at the conservatory and is of all participants most involved with music in his daily life, had an MBEA score of 73.33%; only three participants scored lower. This observation might lead us to expect that the shortened version of a test for amusia does not provide us with an appropriate test observation of the participants' musicality. Therefore, a Pearson correlation coefficient was computed to assess the relationship between the scores on the MBEA and the self-reported musicality ('not musical', 'a bit musical', and 'musical', see 2.3). Surprisingly, a negative correlation (r=-0.299, n=40, p=0.061) was found. Therefore, a one-way between subjects ANOVA was conducted with the AXB score as dependent variable and the self-reported musicality as fixed factor, to compare the effect of the self-reported musicality on the AXB score in the three conditions ('not musical', 'a bit musical', and 'musical'). There was a significant effect of the self-reported musicality on the AXB score, meaning that the three groups differed significantly ($F(2,37)=8.06$; p=0.001). A post hoc comparison using the Tukey HSD test indicated that the mean AXB score for the 'musical' condition (M=84.43, 95% CI [78.72, 90.15]) was significantly higher than the mean AXB score for the 'not musical' condition (M=70.46, 95% CI [64.40, 76.52], p=0.005). However, comparisons between the 'a bit musical' condition (M=66.54, 95% CI [56.63, 76.43]) and the other two groups were not statistically significant at p< .05. This analysis seems to indicate that the self-reported musicality does have an effect on discrimination abilities, that is, that people who engage more in music in their daily life have better discrimination abilities.

4.2.2 Differences in the training in the speech condition and the music condition

Another possible explanation for the results has to do with differences in the training for the speech condition and the music condition. The vowels in the music condition were longer; therefore, there were fewer occurrences of the target vowels in the music condition. It could be that it is not the total duration of exposure to the vowels, but the number of occurrences of the vowels that is most important to generate a learning effect. Previous research has indicated that a training phase with vowel durations shorter than 200 ms can already have an effect, see for example Escudero et al. (2011), with stimuli durations of 93, 140 and 216 ms, and Wanrooij and Boersma (2013) with stimuli durations of 140 ms. These studies all use vowels that are even shorter than the vowels in the training of the speech condition of this thesis. However, since more vowels occurred in the speech condition, there were also more pauses, because after each vowel a pause of 0.1 ms was inserted. This means that even though the participants in the speech condition were exposed to more vowel occurrences, they were also less exposed to vowel sound; that is, the total duration of actual vowel sound was shorter.

Furthermore, the vowels in the AXB test had a length of 245 ms. The training vowels in the speech condition had durations between 200 and 250 ms, whereas the vowels in the music condition had durations between 250 and 750 ms. It might be the case that the participants in the speech condition were better trained for the vowels in the AXB test, because they had been trained with vowels of almost the same duration.

A second difference between the conditions regarding the vowels in the training can be noted when looking at the pitch contours. The vowels in the speech condition all have either ascending or falling pitch, whereas the vowels in the music condition have level pitch. Trainor and Desjardins (2002) found a significant main effect for the advantage of falling pitch contours over steady pitch contours on discrimination ability of vowels[22] for infants.

[22] /i/ and /ɪ/.

This suggests the possibility that the pitch movement in the speech condition aided discrimination ability in the AXB task. However, even though pitch movement might actually have had an effect on discrimination ability, and therefore we might have wanted to control for pitch movement, we could argue against this necessity. That is, it is intrinsic to speech that it has pitch movement in short vowels, and intrinsic to music that it does not have pitch movement in short tones/vowels, because melodies are built up of notes with constant pitch. A melody is only perceived as a melody because of the succession of these constant pitches.

In sum, shorter vowel duration in the speech condition might have aided the participants more than the longer vowels in the music condition for two reasons: because the vowels were shorter, there were more occurrences, and the vowels in the AXB task had the same short duration, which might have made the AXB task easier for the participants in the speech condition. Also, participants may have benefitted from the pitch movement of the vowels in the speech condition as opposed to the steady pitch of the vowels in the music condition.

4.2.3 The peaks in the bimodal distribution

The non-effect of the training could have also been caused by the difference in the F2 difference between the peaks in the bimodal distribution in the training and the F2 difference of the vowel pairs in the AXB task. The bimodal distribution in the training was based on the Slovenian vowels, meaning that the two peaks of the distribution had an F2 difference of 188 Hz. In contrast, the vowels in the AXB task had an F2 difference varying from 88 Hz to 144 Hz. This might explain why the training did not have an effect on the AXB score.

4.2.4 Attention task

Other than the possibility that the vowels in the training phase might have influenced the results on the AXB task, another possible explanation for the fact that no difference between the conditions was found can be taken into consideration. During the training phase, the participants were asked to count the occurrences of /a/, which was an attention consuming task, since the speech condition contained 87 occurrences of /a/, and the music condition contained 48 occurrences of /a/. This task was given in order to make sure that the participants paid attention to the sound, but it might have had a counterproductive effect, namely that the participants were focussing so much on the /a/ in the sound, that they did not register the occurrences of /ɘ/ and /ɜ/. In psychology, the visual phenomenon *inattentional blindness* has been studied intensively. The phenomenon was first studied by Simons and

Chabris (1999), who showed participants a video[23] in which two groups were passing a ball. The participants were then asked to count the number of times that the group with the white shirts passed the ball. During the video, a person in a gorilla suit entered the room, drummed on his/her chest in the middle of the group, and walked away. 50% of the participants did not notice the gorilla, because they focused so much on the counting exercise. Some research has mentioned the possibility that the same effect arises in auditory processing (e.g. Mack 2003 and Harding et al. 2007). Harding et al. suggest that, like in vision, auditory processing starts with the perception of the gist of the signal, and that the relevant components are only perceived afterwards, when being paid attention to. They point out that we can hear music playing or people talking without noticing what instruments are played on, or what people are talking about, meaning that we do automatically perceive the gist of a situation, but not the details. Also, *change deafness* has been reported, for example Vitevitch (2003) found that people participating in a lexical decision task[24] fail to notice that the voice that reads the words changes into another voice. This indicates that even fairly obvious changes can be unnoticed when unattended to.

In other words, it might be the case that the participants paid so much attention to the attention task that they failed to register the target vowels, and therefore entered the AXB task untrained. This would explain the fact that even though the conditions differed considerably, there was no difference in the scores on the AXB task. Unfortunately, there was no control condition to give a decisive answer about the possible non-effect of the training.

4.3 Further observations

In addition to the interpretation of the results, some other observations have been made. First, there were remarkable individual differences. Two participants scored 96,88% correct on the AXB task: they only made one mistake, which means that they were able to hear the difference very well. The lowest scores on the AXB task were 46.88 %, which is below chance level, meaning that the participants that scored so low did not, or barely, hear a difference. Even more surprising are participant 6KG's results in pilot 1. This participant did not even hear a difference between the original Slovenian vowels, which have an F2 difference of 188 Hz. Great individual differences in the ability to discriminate speech sounds have been reported before, for example in Kewley-Port (2001), who tested well-trained and untrained listeners. She concludes that untrained individuals vary greatly in their abilities to

[23] See http://www.youtube.com/watch?v=vJG698U2Mvo

[24] A task in which participants are asked whether or not the presented item (word or nonword) is a word.

discriminate vowels in synthesized speech, since she observed that some untrained listeners showed equal performance as trained listeners after only one hour. Kempe et al. (2012) also investigated individual differences; one of the things they found was an advantage for male participants in non-native speech sound processing. There was an unequal spreading of men in the conditions in the present study: in the speech condition seven men were tested, whereas in the music condition only six men were tested. A sex difference could thus have influenced the outcome on the AXB task. However, when comparing the AXB scores grouped for sex in the present study, no significant difference was found ($t(38)=-1{,}1211$, $p=0.234$).

Secondly, many participants reported that during the AXB task, they were often under the impression that sound one (A) and sound three (B) were the same. Since this was never true, it could raise questions about perception during an AXB task. It might be the case that the order in which the sounds are presented influences the way the sounds are perceived.

5. Conclusion

In this thesis I investigated whether the addition of musical features to a distributional learning task can help the learning of an unknown vowel contrast. 40 participants listened to the vowels of an unknown vowel contrast on either intonation contours or melodies, after which they were tested on their ability to discriminate the vowels. A comparison of the mean scores of the two groups showed no significant difference. Thus, the hypothesis 'the addition of musical features increases the learning effect of a training phase with a bimodal distribution of an unknown vowel contrast, and therefore leads to better results in the discrimination of this contrast' was not confirmed in this research. Also, the additional hypothesis 'participants who score high on the *Montreal Battery of Evaluation of Amusia* will benefit more from the musical features in the music condition, and will therefore obtain better results in the discrimination of the vowel contrast' was not confirmed. In the discussion, however, some possible reasons for the results have been given. More research is necessary to obtain a decisive answer to the hypotheses. For example, it would be interesting to repeat the final experiment with some adaptations:

(1) Faster melodies in the music condition, which would result in more similar vowel durations in the two conditions.

(2) Addition of one speech condition and one music condition in which the attention task would be substituted for the instruction 'It is very important that you listen closely to the sounds'.

(3) Addition of a control condition, to control for the possibility that the training has no effect.

(4) Substitution of the MBEA task for a task that is specifically designed to test musical aptitude, e.g. Wing's *standardized test of musical intelligence* or Seashore's *measures of musical talent*.

This kind of experiment would address two questions. First, it would repeat the hypothesis investigated in this research. Second, it would assess an interesting methodological question, namely whether an attention task as the one applied in this thesis can cause a non-effect of a distributional learning task. However, taking out the attention task would obviously result in a methodological problem, since it can be assumed that participants prefer listening to music over listening to speech. This could cause the participants to listen better to the sounds in the training phase of the music condition than to those in the speech condition. This would probably result in a better performance on the AXB task, which would not be due to the addition of musical features, but to the fact that the participants listened better to the sounds in the music condition. On the other hand, it could be argued that the fact that people like to listen to music is an intrinsic component of music, so it would still be due to the addition of 'music'.

In addition, the question has come up whether it could be the case that it is not the total duration of exposure to the vowels, but the number of occurrences of the vowels that is associated with the greatest learning effect. Training the participants in two different training conditions with the same total duration could test this, where one condition contained many short vowels and the other less longer vowels.

Finally, since so many participants reported that they perceived the first and the third sound in the AXB task as being most similar, it could be investigated whether the order in which the sounds are presented influences the way sounds are perceived.

6. References

Adank, P., Van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical society of America*, *116*(3), 1729-1738.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese/r/and English/l/and/r. *Journal of Phonetics*, *32*(2), 233-250.

Boersma, Paul & Weenink, David (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.49, retrieved 13 May 2013 from http://www.praat.org/

Bornstien, M. H. (2010). Information processing (habituation) in infancy and stability in cognitive development. *Human Development*, *32*(3-4), 129-136.

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, *23*(10), 2791-2803.

Chan, A. S., Ho, Y. C., & Cheung, M. C. (1998). Music training improves verbal memory. *Nature*, *396*(6707), 128-128.

Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., & Colombo, L. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study. *Brain research 1455*: 75-89.

Dexter, E.S. & Omwake, K.T. (1934). The relation between pitch discrimination and accent in modern languages. *Journal of Applied Psychology, 18*, 267-271.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, *37*(4), 452-465.

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, *130*(4), EL206-EL212.

Flege, J.E. (2003). Assessing constraints on second-language segmental production and perception. In Meyer, A. & Schiller, N. (Eds.) *Phonetics and phonology in language* *comprehension and production: Differences and similarities*. Berlin: Mouton de Gruyter, 319-355.

Gussenhoven, C., Rietveld, T., Kerkhoff, J.. Terken, J. (2003) ToDI second edition. Website: http://todi.let.kun.nl/ToDI/home.htm, seen on April 16, 2013.

Harding, S., Cooke, M., & König, P. (2007). Auditory gist perception: an alternative to attentional selection of auditory streams?. In *Attention in Cognitive Systems. Theories and Systems from an Interdisciplinary Viewpoint*. Springer Berlin Heidelberg, 399-416.

Harrison, N. (1978). *A study to find out if a correlation exists between the study of music and aural-oral ability in French* (Doctoral dissertation, Simon Fraser University.).

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569-1579.

Hogan, D. E., & Huesman, T. (2008). Music training and semantic clustering in college students. *The Journal of genetic psychology*, *169*(4), 322-331.

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *Neuroimage*, *47*(2), 735-744.

Karimer, L. (1984). Can southeast Asian students learn to discriminate between English phonemes more quickly with the aid of music and rhythm?. Language Key to Learning (5), 41-48.

Kempe, V., Thoresen, J. C., Kirk, N. W., Schaeffler, F., & Brooks, P. J. (2012). Individual differences in the discrimination of novel speech sounds: effects of sex, temporal processing, musical and cognitive abilities. *PloS one*, *7*(11), e48623.

Kewley-Port, D. (2001). Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training. *The Journal of the Acoustical Society of America*, *110*(4), 2141-2155.

Kolinsky, R., Lidji, P., Peretz, I., Besson, M., & Morais, J. (2009). Processing interactions between phonology and melody: Vowels sing but consonants speak. *Cognition*, *112*(1), 1-20.

Lebedeva, G. C., & Kuhl, P. K. (2010). Sing that tune: Infants' perception of melody and lyrics and the facilitation of phonetic recognition in songs. *Infant Behavior and Development*, *33*(4), 419-430.

Lerdahl, F., & Jackendoff, R. (1983) *A Generative Theory of Tonal Music*. London: MIT Press.

Lowe, A. S. (1998). L'enseignement de la musique et de la langue seconde: pistes d'intégration et conséquences sur les apprentissages. *Canadian Modern Language Review/La Revue canadienne des langues vivantes*, *54*(2), 223-238.

Mack, A. (2003). Inattentional blindness: Looking without seeing. *Current Directions in Psychological Science*, *12*(5), 180-184.

Marin, M. M. (2009). Effects of early musical training on musical and linguistic syntactic abilities. *Annals of the New York Academy of Sciences*, *1169*(1), 187-190.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In *Proceedings of the 24th Annual Boston University Conference on Language Development, 2.* Somerville, MA: Cascadilla Press, 522-533.

Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us. In *Proceedings of the 25th annual Boston University conference on language development*. Cascadilla Press, 480-490.

van der Meulen, I., van de Sandt-Koenderman, M. E., & Ribbers, G. M. (2012). Melodic intonation therapy: present controversies and future opportunities. *Archives of physical medicine and rehabilitation*, *93*(1), S46-S52.

Mermelstein, P. (1978). Difference limens for formant frequencies of steady-state and consonant-bound vowels. *The Journal of the Acoustical Society of America*, *63*, 572-580.

Mora, C. F. (2000). Foreign language acquisition and melody singing. *ELT journal*, *54*(2), 146-152.

Nardo, D. & Reiterer, S.M. (2009). Musicality and phonetic language aptitude. In Dogil, G. & Reiterer, S. M. (Eds.) 2009 *Language talent and brain activity.* Mouton de Gruyter, Berlijn, 213-255.

Nooteboom, S. (1997). The prosody of speech: melody and rhythm. Hardcastle, W.J. & Laver, J. (Eds.) *The Handbook of Phonetic Sciences*, (5), 640-673.

Norton, A., Zipse, L., Marchina, S., & Schlaug, G. (2009). Melodic intonation therapy. *Annals of the New York Academy of Sciences*, *1169*(1), 431-436.

Pastuszek-Lipińska, B. (2008). Influence of music education on second language acquisition. *Journal of the Acoustical Society of America*, *123*(5), 3737, 8781-8786.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. *Annals of the New York Academy of Sciences*, *999*(1), 58-75.

Piro, J. M., & Ortiz, C. (2009). The effect of piano lessons on the vocabulary and verbal sequencing skills of primary grade students. *Psychology of Music*,*37*(3), 325-347.

Repp, B. H., (1991) Some cognitive and perceptual aspects of speech and music. In Sundberg, J., Nord, L., & Carlson, R. (Eds.) 1991, 257-268.

Risset, J. C. (1991) Speech and music combined: an overview. In Sundberg, J., Nord, L., & Carlson, R. (Eds.) 1991, 368-379.

Saffran, J. R. (2003). Musical learning and language development. *Annals of the New York Academy of Sciences*, *999*(1), 397-401.

Sarroff, A. (2009). Musicians and Non-Musicians: Anatomical Differences in the Human Brain. Course instruction Psychology of Music Spring Semester, 2009, New York University.

Schellenberg, E. G., & Trehub, S. E. (1999). Redundancy, conventionality, and the discrimination of tone sequences: A developmental perspective. *Journal of Experimental Child Psychology*, *74*, 107-127.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*(5198), 699-701.

Schlaug, G., Norton, A., Overy, K., & Winner, E. (2005). Effects of music training on the child's brain and cognitive development. *Annals of the New York Academy of Sciences*, *1060*(1), 219-230.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, *106*(2), 975-983.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception-London*, *28*(9), 1059-1074.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter?. *Psychological Science*, *17*(8), 675-681.

Steenwijk, H. (1992). *The Slovene dialect of Resia, San Giorgia.* Proefschrift Rijksuniversiteit Leiden, 22 oktober 1992. Amsterdam: Atlanta, GA.

Schellenberg, E. G., & Trehub, S. E. (1999). Redundancy, conventionality, and the discrimination of tone sequences: A developmental perspective. *Journal of Experimental Child Psychology*, *74*, 107-127.

Sundberg, J., Nord, L., & Carlson, R. (Eds.) (1991). *Music, Language, Speech and Brain: Proceedings of an International Symposium at the Wenner-Gren Center, Stockholm, 5-8 September 1990* (Vol. 59). Basingstoke: Macmillan Press.

Sundberg, J. (1991) Summary. In Sundberg, J., Nord, L., & Carlson, R. (Eds.) 1991, 441-446.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*(2), 335-340.

Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience*, *6*(7), 669-673.

Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology. Human Perception and Performance*, *29*(2), 333-342.

Wanrooij, K., & Boersma, P. (2013). Distributional training of speech sounds can be done with continuous distributions. *The Journal of the Acoustical Society of America*, *133*(5), EL398-EL404.

Welker, C. (2006). Tutorial for self study: basics of phonetics and how to use Praat. Website: http://www.hum.uu.nl/uilots/lab/courseware/phonetics/index.html, seen on June 2, 2013.

7. Appendices

7.1 Scripts

7.1.1 Script used to synthesize the target vowels (adapted from a script by K. Chládková)

```
startPitch   =   210
maxPitch     =   30
minPitch     =   40
openPhase    =   0.75
collisionPhase = 35
spectralTilt   = 2.5
aspirationAmplitude =    0.5
breathinessAmplitudeStart  =     0.55
breathinessAmplitudeEnd =0.50
minVoiceAmpl = 63
maxVoiceAmpl = 72

f1Bandwidth   =  250
f2Bandwidth   =  250
f3Bandwidth   =  250
f4Bandwidth   =  200
f5Bandwidth   =  200
upBandwidth   =  12.5

## the three formants below were varied to create 45 /ɜ/-vowels and 45 /ɘ/-vowels
formant1 = hertzToErb(558)
formant2 = hertzToErb(1709)
formant3 = hertzToErb(3146)

## choose whether you want short (180 ms = 5.2) or long vowel (330 ms = 5.8), or medium (245 ms = 5.5)

duration = 5.5

fric=0
vowel=1

call makeVowel

procedure makeVowel
    #Change to hertz values.
    f1base=erbToHertz(formant1)
    f2base=erbToHertz(formant2)
    f3base=erbToHertz(formant3)
    f1 = f1base
    f2 = f2base
    f3 = f3base
    f1end=f1
    f2end=f2
    f3end=f3

    dur=e^duration/1000

    collPhase=formant1/collisionPhase

    #Set parameters I might need later
    nrOfFormants=20
    nrOfNasalFormants=0
    nrOfNasalAntiFormants = 0
    nrOfTrachealFormants = 0
    nrOfTrachealAntiFormants = 0
```

```
    nrOfFricationFormants = 0
    nrOfDeltaFormants = 1

    startingF0 = startPitch
    startF0 = hertzToMel(startingF0)
    maxF0 = startF0 +maxPitch
    minF0 = startF0 -minPitch


    #Set duration parameters that might be different in SVS script

    durVowel = dur
    startVowel=0
    endVowel=durVowel
    durTotal=durVowel

    # Add some extra formants to get a flatter spectrum.
    f4 = max (3850, f3 + 400)
    f5 = max (4950, f4 + 600)
    f4end=f4
    f5end=f5
    for highFormant from 6 to 20
            lowerByOne = 'highFormant'-1
            f'highFormant' = f'lowerByOne' + 1000
            f'highFormant'end = f'highFormant'
    endfor

    #Create the Grid, with the duration that's currently being used
    grid = Create KlattGrid... vowel 0 durTotal nrOfFormants nrOfNasalFormants nrOfNasalAntiFormants
… nrOfFricationFormants
    ... nrOfTrachealFormants nrOfTrachealAntiFormants  nrOfDeltaFormants

    #Call the procedures (see below)

    call pitchPhonation
    call voiceQuality
    call formant

    #Make the sound
    nowarn To Sound
    nowarn To Sound (special)... 0 dur 44100 y y y y y y "Powers in tiers" y y y Cascade 1 nrOfFormants 1
; … nrOfNasalFormants 1 nrOfNasalAntiFormants
    ; ... 1 nrOfTrachealFormants 1 nrOfTrachealAntiFormants 0 0  0 0 1 nrOfFricationFormants n
    ; ... y n y

    #Window en scaling
    Fade in... All 0 0.005 n
    Fade out... All dur -0.005 n
    Scale peak... 0.99
    select grid
    Remove
    select Sound vowel
endproc


### PITCH, IN sinoid CURVE, PHONATION AMPLITUDE IN linear decrease
procedure pitchPhonation
    Add pitch point... startVowel+0.15*(endVowel-startVowel) melToHertz(maxF0)
    Add pitch point... startVowel melToHertz(startF0)
    Add pitch point... endVowel melToHertz(minF0)

    Add voicing amplitude point... startVowel maxVoiceAmpl
    Add voicing amplitude point... endVowel minVoiceAmpl
    Add breathiness amplitude point... startVowel maxVoiceAmpl*breathinessAmplitudeStart
    Add breathiness amplitude point... endVowel minVoiceAmpl*breathinessAmplitudeEnd

endproc
```

```
procedure voiceQuality
    ### FLUTTER
    Add flutter point... startVowel 0.15

    ### OPEN PHASE
    Add open phase point... startVowel openPhase

    ### FLOW FUNCTION
    Add power1 point... startVowel 3
    Add power2 point... startVowel 4

    ;### SPECTRAL TILT
    Add spectral tilt point... startVowel spectralTilt

    ### COLLISION PHASE
    Add collision phase point... startVowel collPhase

    ### DELTA
    Add delta formant bandwidth point... 1 startVowel 100
endproc

### ADD FORMANTS
procedure formant
    for f from 1 to 20
            freq = f'f'
            endfreq = f'f'end
            ;bw=sqrt(80^2+(freq/20)^2)
            if f=1
                    bw=f1Bandwidth*formant1/formant2
            elsif f<6
                    bw = f'f'Bandwidth
            else
                    bw=freq/upBandwidth
            endif

            Add oral formant frequency point... f startVowel freq
            Add oral formant bandwidth point... f startVowel bw

            Add oral formant frequency point... f endVowel endfreq
            Add oral formant bandwidth point... f endVowel bw
    endfor

endproc
```

7.1.2 Scripts used to create the strings for the training (includes the script from 7.1.1)

```
filename$ = "into1"

Create Sound from formula... startingsilence 0.5 0 1 44100 0
Save as WAV file... ~/Desktop/'filename$'.wav

select Sound startingsilence
Remove

for i from 1 to 13
klinker = i+1
klinker = randomInteger(1,12)
    if klinker <=2
    klinker$= "a"
    elsif klinker <= 4
    klinker$="i"
    elsif klinker <= 6
    klinker$="u"
    elsif klinker <= 9
    klinker$="e"
```

```
        elsif klinker <= 12
        klinker$="o"
        endif
printline klinker'i' = 'klinker$'
vowel'i'$ = klinker$
endfor

call speel'vowel1$' 258 -20 200 0.1
call speel'vowel2$' 250 95 230 0.1
call speel'vowel3$' 327 -68 200 0.1
call speel'vowel4$' 270 -36 250 0.1
call speel'vowel5$' 257 -24 200 0.1
call speel'vowel6$' 240 -3 230 0.1
call speel'vowel7$' 237 16 200 0.1
call speel'vowel8$' 253 54 250 0.1
call speel'vowel9$' 318 -100 200 0.1
call speel'vowel10$' 223 -10 200 0.1
call speel'vowel11$' 300 -100 200 0.1
call speel'vowel12$' 209 -5 230 0.1
call speel'vowel13$' 206 -90 250 0.7

procedure speela beginaf0 changeaf0 adur apause

        startPitch   =   beginaf0
        pitchChange = changeaf0
        openPhase    =   0.75
        collisionPhase = 35
        spectralTilt   = 2.5
        aspirationAmplitude =    0.5
        breathinessAmplitudeStart  =     0.55
        breathinessAmplitudeEnd =0.50
        minVoiceAmpl = 63
        maxVoiceAmpl = 72

        f1Bandwidth   =  250
        f2Bandwidth   =  250
        f3Bandwidth   =  250
        f4Bandwidth   =  200
        f5Bandwidth   =  200
        upBandwidth   =  12.5

        formant1Hz = randomGauss(912,15)
        formant1 = hertzToErb(formant1Hz)
        formant2 = hertzToErb(1572)
        formant3 = hertzToErb(2852)

      ## choose whether you want short (180 ms = 5.2) or long vowel (330 ms = 5.8), or medium (245 ms = 5.5)

        logdur = ln(adur)
        duration = logdur

        fric=0
        vowel=1

        call makeVowel
        Play
        Append to existing sound file... ~/Desktop/'filename$'.wav
        Remove

        do ("Create Sound as pure tone...", "silence", 1, 0, apause, 44100, 1, 1e-05, 0.0001, 0.0001)
        Play
        Append to existing sound file... ~/Desktop/'filename$'.wav
        Remove
endproc

procedure speelu beginuf0 changeuf0 udur upause
```

```
     startPitch    =    beginuf0
     pitchChange = changeuf0
     openPhase     =    0.75
     collisionPhase = 35
     spectralTilt    = 2.5
     aspirationAmplitude = 0.5
     breathinessAmplitudeStart  =     0.6
     breathinessAmplitudeEnd =0.55
     minVoiceAmpl = 63
     maxVoiceAmpl = 72

     f1Bandwidth   =  250
     f2Bandwidth   =  250
     f3Bandwidth   =  250
     f4Bandwidth   =  200
     f5Bandwidth   =  200
     upBandwidth    =  12.5

     formant1Hz = randomGauss(286,15)
     formant1 = hertzToErb(formant1Hz)
     formant2 = hertzToErb(938)
     formant3 = hertzToErb(2736)

    ## choose whether you want short (180 ms = 5.2) or long vowel (330 ms = 5.8), or medium (245 ms = 5.5)

     logdur = ln(udur)
     duration = logdur

     fric=0
     vowel=1

     call makeVowel
     Play
     Append to existing sound file... ~/Desktop/'filename$'.wav
     Remove

     do ("Create Sound as pure tone...", "silence", 1, 0, upause, 44100, 1, 1e-05, 0.0001, 0.0001)
     Play
     Append to existing sound file... ~/Desktop/'filename$'.wav
     Remove
endproc

procedure speeli beginif0 changeif0 idur ipause

     startPitch    =    beginif0
     pitchChange = changeif0
     openPhase     =    0.75
     collisionPhase = 35
     spectralTilt    = 2.5
     aspirationAmplitude =    0.5
     breathinessAmplitudeStart  =     0.6
     breathinessAmplitudeEnd =0.55
     minVoiceAmpl = 63
     maxVoiceAmpl = 72

     f1Bandwidth   =  250
     f2Bandwidth   =  250
     f3Bandwidth   =  250
     f4Bandwidth   =  200
     f5Bandwidth   =  200
     upBandwidth    =  12.5

     formant1Hz = randomGauss(294,15)
     formant1 = hertzToErb(formant1Hz)
     formant2 = hertzToErb(2524)
     formant3 = hertzToErb(2911)
```

```
    ## choose whether you want short (180 ms = 5.2) or long vowel (330 ms = 5.8), or medium (245 ms = 5.5)

    logdur = ln(idur)
    duration = logdur

    fric=0
    vowel=1

    call makeVowel
    Play
    Append to existing sound file... ~/Desktop/'filename$'.wav
    Remove

    do ("Create Sound as pure tone...", "silence", 1, 0, ipause, 44100, 1, 1e-05, 0.0001, 0.0001)
    Play
    Append to existing sound file... ~/Desktop/'filename$'.wav
    Remove
endproc

procedure speele beginef0 changeef0 edur epause

    startPitch   =   beginef0
    pitchChange = changeef0
    openPhase    =   0.75
    collisionPhase = 35
    spectralTilt   = 2.5
    aspirationAmplitude =    0.5
    breathinessAmplitudeStart  =     0.55
    breathinessAmplitudeEnd =0.50
    minVoiceAmpl = 63
    maxVoiceAmpl = 72

    f1Bandwidth   =  250
    f2Bandwidth   =  250
    f3Bandwidth   =  250
    f4Bandwidth   =  200
    f5Bandwidth   =  200
    upBandwidth   =  12.5

    formant2Hz = randomGauss(1709, 30)
    formant1 = hertzToErb(558)
    formant2 = hertzToErb(formant2Hz)
    formant3 = hertzToErb(3146)

    ## choose whether you want short (180 ms = 5.2) or long vowel (330 ms = 5.8), or medium (245 ms = 5.5)


    logdur = ln(edur)
    duration = logdur

    fric=0
    vowel=1

    call makeVowel
    Play
    Append to existing sound file... ~/Desktop/'filename$'.wav
    Remove

    do ("Create Sound as pure tone...", "silence", 1, 0, epause, 44100, 1, 1e-05, 0.0001, 0.0001)
    Play
    Append to existing sound file... ~/Desktop/'filename$'.wav
    Remove

endproc

procedure speelo beginof0 changeof0 odur opause
```

```
    startPitch   =   beginof0
    pitchChange = changeof0
    openPhase    =   0.75
    collisionPhase = 35
    spectralTilt   = 2.5
    aspirationAmplitude =    0.5
    breathinessAmplitudeStart  =     0.55
    breathinessAmplitudeEnd =0.50
    minVoiceAmpl = 63
    maxVoiceAmpl = 72

    f1Bandwidth   =  250
    f2Bandwidth   =  250
    f3Bandwidth   =  250
    f4Bandwidth   =  200
    f5Bandwidth   =  200
    upBandwidth   =  12.5

    formant2Hz = randomGauss(1521, 30)
    formant1 = hertzToErb(558)
    formant2 = hertzToErb(formant2Hz)
    formant3 = hertzToErb(3146)

   ## choose whether you want short (180 ms = 5.2) or long vowel (330 ms = 5.8), or medium (245 ms = 5.5)

    logdur = ln(odur)
    duration = logdur

    fric=0
    vowel=1

    call makeVowel
    Play
    Append to existing sound file... ~/Desktop/'filename$'.wav
    Remove

    do ("Create Sound as pure tone...", "silence", 1, 0, opause, 44100, 1, 1e-05, 0.0001, 0.0001)
    Play
    Append to existing sound file... ~/Desktop/'filename$'.wav
    Remove

endproc

### PROCEDURE makeVowel
procedure makeVowel
    #Change to hertz values.
    f1base=erbToHertz(formant1)
    f2base=erbToHertz(formant2)
    f3base=erbToHertz(formant3)
    f1 = f1base
    f2 = f2base
    f3 = f3base
    f1end=f1
    f2end=f2
    f3end=f3

    dur=e^duration/1000

    collPhase=formant1/collisionPhase

    #Set parameters I might need later
    nrOfFormants=20
    nrOfNasalFormants=0
    nrOfNasalAntiFormants = 0
    nrOfTrachealFormants = 0
    nrOfTrachealAntiFormants = 0
    nrOfFricationFormants = 0
```

```
    nrOfDeltaFormants = 1

    startF0 = startPitch
    endF0 = startPitch + pitchChange
    #Set duration parameters that might be different in SVS script

    durVowel = dur
    startVowel=0
    endVowel=durVowel
    durTotal=durVowel

    # Add some extra formants to get a flatter spectrum.
    f4 = max (3850, f3 + 400)
    f5 = max (4950, f4 + 600)
    f4end=f4
    f5end=f5
    for highFormant from 6 to 20
            lowerByOne = 'highFormant'-1
            f'highFormant' = f'lowerByOne' + 1000
            f'highFormant'end = f'highFormant'
    endfor

    #Create the Grid, with the duration that's currently being used
            grid = Create KlattGrid... vowel 0 durTotal nrOfFormants nrOfNasalFormants
    … nrOfNasalAntiFormants nrOfFricationFormants
    ... nrOfTrachealFormants nrOfTrachealAntiFormants  nrOfDeltaFormants

    #Call the procedures (see below)

    call pitchPhonation
    call voiceQuality
    call formant

    #Make the sound
    nowarn To Sound
;   nowarn To Sound (special)... 0 dur 44100 y y y y y y "Powers in tiers" y y y Cascade 1 nrOfFormants
    … 1  nrOfNasalFormants 1 nrOfNasalAntiFormants
;   ... 1 nrOfTrachealFormants 1 nrOfTrachealAntiFormants 0 0  0 0 1 nrOfFricationFormants n
;   ... y n y

    #Window en scaling
    Fade in... All 0 0.005 n
    Fade out... All dur -0.005 n
    Scale peak... 0.99
    select grid
    Remove
    select Sound vowel
endproc

### PROCEDURE 1 called by makeVowel (PITCH, IN sinoid CURVE, PHONATION AMPLITUDE IN
linear decrease)
    procedure pitchPhonation
            Add pitch point... startVowel startF0
            Add pitch point... endVowel endF0

            Add voicing amplitude point... startVowel maxVoiceAmpl
            Add voicing amplitude point... endVowel minVoiceAmpl
            Add breathiness amplitude point... startVowel maxVoiceAmpl*breathinessAmplitudeStart
            Add breathiness amplitude point... endVowel minVoiceAmpl*breathinessAmplitudeEnd

    endproc

### PROCEDURE 2 called by makeVowel
    procedure voiceQuality
            Add flutter point... startVowel 0.15
            Add open phase point... startVowel openPhase
            Add power1 point... startVowel 3
```

```
			Add power2 point... startVowel 4
			Add spectral tilt point... startVowel spectralTilt
			Add collision phase point... startVowel collPhase
			Add delta formant bandwidth point... 1 startVowel 100
	endproc

### PROCEDURE 3 called by makeVowel
	procedure formant
			for f from 1 to 20
					freq = f'f'
					endfreq = f'f'end
					;bw=sqrt(80^2+(freq/20)^2)
					if f=1
							bw=f1Bandwidth*formant1/formant2
					elsif f<6
							bw = f'f'Bandwidth
					else
							bw=freq/upBandwidth
					endif

					Add oral formant frequency point... f startVowel freq
					Add oral formant bandwidth point... f startVowel bw

					Add oral formant frequency point... f endVowel endfreq
					Add oral formant bandwidth point... f endVowel bw
			endfor
	endproc
```

To create the other 7 strings, the first 40 lines of this script were replaced.

For the second intonation contour the first 40 lines were replaced by:

```
filename$ = "into2"

Create Sound from formula... startingsilence 0.5 0 1 44100 0
Save as WAV file... ~/Desktop/'filename$'.wav

select Sound startingsilence
Remove

for i from 1 to 10
klinker = i+1
	klinker = randomInteger(1,12)
	if klinker <=2
	klinker$= "a"
	elsif klinker <= 4
	klinker$="i"
	elsif klinker <= 6
	klinker$="u"
	elsif klinker <= 9
	klinker$="e"
	elsif klinker <= 12
	klinker$="o"
	endif
printline klinker'i' = 'klinker$'
vowel'i'$ = klinker$
endfor


call speel'vowel1$' 311 -11 200 0.1
call speel'vowel2$' 274 24 230 0.1
call speel'vowel3$' 298 -11 200 0.1
call speel'vowel4$' 308 -58 230 0.1
call speel'vowel5$' 246 8 200 0.1
```

```
call speel'vowel6$' 246 94 250 0.1
call speel'vowel7$' 370 -28 250 0.1
call speel'vowel8$' 333 -46 200 0.1
call speel'vowel9$' 282 -37 230 0.1
call speel'vowel10$' 245 -9 250 0.7
```

For the third intonation contour the first 40 lines were replaced by:

```
filename$ = "into3"

Create Sound from formula... startingsilence 0.5 0 1 44100 0
Save as WAV file... ~/Desktop/'filename$'.wav

select Sound startingsilence
Remove

for i from 1 to 20
klinker = i+1
    klinker = randomInteger(1,12)
    if klinker <=2
    klinker$= "a"
    elsif klinker <= 4
    klinker$="i"
    elsif klinker <= 6
    klinker$="u"
    elsif klinker <= 9
    klinker$="e"
    elsif klinker <= 12
    klinker$="o"
    endif
printline klinker'i' = 'klinker$'
vowel'i'$ = klinker$
endfor


call speel'vowel1$' 284 -63 200 0.1
call speel'vowel2$' 251 16 200 0.1
call speel'vowel3$' 250 22 200 0.1
call speel'vowel4$' 272 -28 230 0.1
call speel'vowel5$' 240 -23 200 0.1
call speel'vowel6$' 218 100 250 0.1
call speel'vowel7$' 281 -42 200 0.1
call speel'vowel8$' 247 -25 250 0.1
call speel'vowel9$' 220 33 230 0.6

call speel'vowel10$' 237 23 200 0.1
call speel'vowel11$' 214 75 250 0.1
call speel'vowel12$' 285 -44 220 0.1
call speel'vowel13$' 248 -27 200 0.1
call speel'vowel14$' 220 12 200 0.1
call speel'vowel15$' 263 95 230 0.1
call speel'vowel16$' 277 -53 210 0.1
call speel'vowel17$' 224 -6 210 0.1
call speel'vowel18$' 218 21 220 0.1
call speel'vowel19$' 282 23 250 0.1
call speel'vowel20$' 205 -15 250 0.7
```

For the fourth intonation contour the first 40 lines were replaced by:

```
filename$ = "into4"

Create Sound from formula... startingsilence 0.5 0 1 44100 0
Save as WAV file... ~/Desktop/'filename$'.wav
```

```
select Sound startingsilence
Remove

for i from 1 to 12
klinker = i+1
     klinker = randomInteger(1,12)
     if klinker <=2
     klinker$= "a"
     elsif klinker <= 4
     klinker$="i"
     elsif klinker <= 6
     klinker$="u"
     elsif klinker <= 9
     klinker$="e"
     elsif klinker <= 12
     klinker$="o"
     endif
printline klinker'i' = 'klinker$'
vowel'i'$ = klinker$
endfor


call speel'vowel1$' 291 -78 200 0.1
call speel'vowel2$' 213 27 230 0.1
call speel'vowel3$' 240 -19 200 0.1
call speel'vowel4$' 218 -19 200 0.1
call speel'vowel5$' 215 100 250 0.1
call speel'vowel6$' 315 -55 210 0.1
call speel'vowel7$' 264 -56 250 0.1
call speel'vowel8$' 241 67 220 0.1
call speel'vowel9$' 308 46 200 0.1
call speel'vowel10$' 340 -100 200 0.1
call speel'vowel11$' 240 -26 250 0.1
call speel'vowel12$' 214 -11 230 0.7
```

For the first melody the first 40 lines were replaced by:

```
filename$ = "melo1"

Create Sound from formula... startingsilence 0.5 0 1 44100 0
Save as WAV file... ~/Desktop/'filename$'.wav

select Sound startingsilence
Remove

for i from 1 to 19
klinker = i+1
     klinker = randomInteger(1,12)
     if klinker <=2
     klinker$= "a"
     elsif klinker <= 4
     klinker$="i"
     elsif klinker <= 6
     klinker$="u"
     elsif klinker <= 9
     klinker$="e"
     elsif klinker <= 12
     klinker$="o"
     endif
printline klinker'i' = 'klinker$'
vowel'i'$ = klinker$
endfor

### n stands for ‘note’, the next letter(s) indicate(s) which note it is. From c on there should have been an
## apostrophe, because c is the central c (c’).
```

```
ng = 195.99771799087463
ngis = 207.65234878997256
na = 220
nbes = 233.08188075904496
nb = 246.94165062806206
nc = 261.6255653005986
ncis = 277.1826309768721
nd = 293.6647679174076
ndis = 311.12698372208087
nf = 349.2282314330039

heel = 1000
half = 500
kwart = 250
eindpauze = 0.1

call speel'vowel1$' ng kwart eindpauze
call speel'vowel2$' na kwart eindpauze
call speel'vowel3$' nbes half eindpauze
call speel'vowel4$' na kwart eindpauze
call speel'vowel5$' nbes half eindpauze

call speel'vowel6$' nd kwart eindpauze
call speel'vowel7$' nc kwart eindpauze
call speel'vowel8$' nbes kwart eindpauze
call speel'vowel9$' nc kwart eindpauze
call speel'vowel10$' nd half 0.7
```

For the second melody the first 40 lines were replaced by:

```
filename$ = "melo2"

Create Sound from formula... startingsilence 0.5 0 1 44100 0
Save as WAV file... ~/Desktop/'filename$'.wav

select Sound startingsilence
Remove

for i from 1 to 19
klinker = i+1
    klinker = randomInteger(1,12)
    if klinker <=2
    klinker$= "a"
    elsif klinker <= 4
    klinker$="i"
    elsif klinker <= 6
    klinker$="u"
    elsif klinker <= 9
    klinker$="e"
    elsif klinker <= 12
    klinker$="o"
    endif
printline klinker'i' = 'klinker$'
vowel'i'$ = klinker$
endfor

### n stands for 'note', the next letter(s) indicate(s) which note it is. From c on there should have been an
## apostrophe, because c is the central c (c').


ng = 195.99771799087463
ngis = 207.65234878997256
na = 220
```

```
nbes = 233.08188075904496
nb = 246.94165062806206
nc = 261.6255653005986
ncis = 277.1826309768721
nd = 293.6647679174076
ndis = 311.12698372208087
nf = 349.2282314330039

heel = 1000
half = 500
kwart = 250
eindpauze = 0.1

call speel'vowel1$' nbes half eindpauze
call speel'vowel2$' nc half eindpauze
call speel'vowel3$' nd half eindpauze
call speel'vowel4$' nf kwart eindpauze
call speel'vowel5$' nd kwart eindpauze

call speel'vowel6$' nc kwart eindpauze
call speel'vowel7$' nd half eindpauze
call speel'vowel8$' nc kwart eindpauze
call speel'vowel9$' nbes kwart eindpauze
call speel'vowel10$' nc half eindpauze
call speel'vowel11$' nbes half 0.7
```

For the third melody the first 40 lines were replaced by:

```
filename$ = "melo3"

Create Sound from formula... startingsilence 0.5 0 1 44100 0
Save as WAV file... ~/Desktop/'filename$'.wav

select Sound startingsilence
Remove

for i from 1 to 19
klinker = i+1
    klinker = randomInteger(1,12)
    if klinker <=2
    klinker$= "a"
    elsif klinker <= 4
    klinker$="i"
    elsif klinker <= 6
    klinker$="u"
    elsif klinker <= 9
    klinker$="e"
    elsif klinker <= 12
    klinker$="o"
    endif
printline klinker'i' = 'klinker$'
vowel'i'$ = klinker$
endfor

### n stands for ‘note’, the next letter(s) indicate(s) which note it is. From c on there should have been an
## apostrophe, because c is the central c (c’).


ng = 195.99771799087463
ngis = 207.65234878997256
na = 220
nbes = 233.08188075904496
nb = 246.94165062806206
nc = 261.6255653005986
ncis = 277.1826309768721
```

```
nd = 293.6647679174076
ndis = 311.12698372208087
nf = 349.2282314330039

heel = 1000
driekwart = 750
half = 500
kwart = 250
eindpauze = 0.1


call speel'vowel1$' nb half eindpauze
call speel'vowel2$' ncis kwart eindpauze
call speel'vowel3$' ndis kwart eindpauze
call speel'vowel4$' nb half eindpauze

call speel'vowel5$' ngis half eindpauze
call speel'vowel6$' ncis kwart eindpauze
call speel'vowel7$' nb heel 0.7
```

For the fourth melody the first 40 lines were replaced by:

```
filename$ = "melo4"

Create Sound from formula... startingsilence 0.5 0 1 44100 0
Save as WAV file... ~/Desktop/'filename$'.wav

select Sound startingsilence
Remove

for i from 1 to 19
klinker = i+1
    klinker = randomInteger(1,12)
    if klinker <=2
    klinker$= "a"
    elsif klinker <= 4
    klinker$="i"
    elsif klinker <= 6
    klinker$="u"
    elsif klinker <= 9
    klinker$="e"
    elsif klinker <= 12
    klinker$="o"
    endif
printline klinker'i' = 'klinker$'
vowel'i'$ = klinker$
endfor

### n stands for ‘note’, the next letter(s) indicate(s) which note it is. From c on there should have been an
## apostrophe, because c is the central c (c’).


ng = 195.99771799087463
ngis = 207.65234878997256
na = 220
nbes = 233.08188075904496
nb = 246.94165062806206
nc = 261.6255653005986
ncis = 277.1826309768721
nd = 293.6647679174076
ndis = 311.12698372208087
nf = 349.2282314330039

heel = 1000
driekwart = 750
```

```
half = 500
kwart = 250
eindpauze = 0.1


call speel'vowel1$' ndis half eindpauze
call speel'vowel2$' ncis kwart eindpauze
call speel'vowel3$' nb half eindpauze

call speel'vowel4$' na kwart eindpauze
call speel'vowel5$' ngis kwart eindpauze
call speel'vowel6$' na kwart eindpauze

call speel'vowel7$' nb half eindpauze
call speel'vowel8$' ncis kwart eindpauze
call speel'vowel9$' nb kwart eindpauze
call speel'vowel10$' ngis kwart 0.7
```

7.1.3 Script used to create the training for the two conditions

Script for the speech condition:

```
Create Sound from formula... startingsilence 0.5 0 0.5 44100 0
Save as WAV file... ~/Desktop/GeluidenConditie1.wav

for i to 40
    nummer = randomInteger(1,4)
    execute into'nummer'.praat
    do ("Read from file...", "/Users/giselagovaart/Desktop/into'nummer'.wav")
endfor

select all
do ("Concatenate")
    Append to existing sound file... ~/Desktop/GeluidenConditie1.wav
```

Script for the music condition:

```
Create Sound from formula... startingsilence 0.5 0 0.5 44100 0
Save as WAV file... ~/Desktop/GeluidenConditie2.wav

for i to 40
    nummer = randomInteger(1,4)
    execute melo'nummer'.praat
    do ("Read from file...", "/Users/giselagovaart/Desktop/melo'nummer'.wav")
endfor

select all
do ("Concatenate")
    Append to existing sound file... ~/Desktop/GeluidenConditie2.wav
```

7.1.4 Experiment script for pilot 1

```
"ooTextFile"
"ExperimentMFC 6"
blankWhilePlaying? <no>
stimuliAreSounds? <yes>
"stimuli/" ".wav"
carrier "" ""
initial silence 0.8
inter-stimulus interval 0.4
```

final silence 0.8

180 stimuli

```
"1f1,1f1,2f1"       ""
"2f1,2f1,1f1"       ""
"1f1,2f1,2f1"       ""
"2f1,1f1,1f1"       ""

"1f2,1f2,2f2"       ""
"2f2,2f2,1f2"       ""
"1f2,2f2,2f2"       ""
"2f2,1f2,1f2"       ""

"1f3,1f2,2f2"       ""
"2f2,2f2,1f2"       ""
"1f2,2f2,2f2"       ""
"2f2,1f2,1f2"       ""

"1f4,1f4,2f4"       ""
"2f4,2f4,1f4"       ""
"1f4,2f4,2f4"       ""
"2f4,1f4,1f4"       ""

"1f5,1f5,2f5"       ""
"2f5,2f5,1f5"       ""
"1f5,2f5,2f5"       ""
"2f5,1f5,1f5"       ""

"1f6,1f6,2f6"       ""
"2f6,2f6,1f6"       ""
"1f6,2f6,2f6"       ""
"2f6,1f6,1f6"       ""

"1f7,1f7,2f7"       ""
"2f7,2f7,1f7"       ""
"1f7,2f7,2f7"       ""
"2f7,1f7,1f7"       ""

"1f8,1f8,2f8"       ""
"2f8,2f8,1f8"       ""
"1f8,2f8,2f8"       ""
"2f8,1f8,1f8"       ""

"1f9,1f9,2f9"       ""
"2f9,2f9,1f9"       ""
"1f9,2f9,2f9"       ""
"2f9,1f9,1f9"       ""

"1f10,1f10,2f10"    ""
"2f10,2f10,1f10"    ""
"1f10,2f10,2f10"    ""
"2f10,1f10,1f10"    ""

"1f11,1f11,2f11"    ""
"2f11,2f11,1f11"    ""
"1f11,2f11,2f11"    ""
"2f11,1f11,1f11"    ""

"1f12,1f12,2f12"    ""
"2f12,2f12,1f12"    ""
"1f12,2f12,2f12"    ""
"2f12,1f12,1f12"    ""

"1f13,1f13,2f13"    ""
"2f13,2f13,1f13"    ""
"1f13,2f13,2f13"    ""
"2f13,1f13,1f13"    ""
```

"1f14,1f14,2f14" ""
"2f14,2f14,1f14" ""
"1f14,2f14,2f14" ""
"2f14,1f14,1f14" ""

"1f15,1f15,2f15" ""
"2f15,2f15,1f15" ""
"1f15,2f15,2f15" ""
"2f15,1f15,1f15" ""

"1f16,1f16,2f16" ""
"2f16,2f16,1f16" ""
"1f16,2f16,2f16" ""
"2f16,1f16,1f16" ""

"1f17,1f17,2f17" ""
"2f17,2f17,1f17" ""
"1f17,2f17,2f17" ""
"2f17,1f17,1f17" ""

"1f18,1f18,2f18" ""
"2f18,2f18,1f18" ""
"1f18,2f18,2f18" ""
"2f18,1f18,1f18" ""

"1f19,1f19,2f19" ""
"2f19,2f19,1f19" ""
"1f19,2f19,2f19" ""
"2f19,1f19,1f19" ""

"1f20,1f20,2f20" ""
"2f20,2f20,1f20" ""
"1f20,2f20,2f20" ""
"2f20,1f20,1f20" ""

"1f21,1f21,2f21" ""
"2f21,2f21,1f21" ""
"1f21,2f21,2f21" ""
"2f21,1f21,1f21" ""

"1f22,1f22,2f22" ""
"2f22,2f22,1f22" ""
"1f22,2f22,2f22" ""
"2f22,1f22,1f22" ""

"1f23,1f23,2f23" ""
"2f23,2f23,1f23" ""
"1f23,2f23,2f23" ""
"2f23,1f23,1f23" ""

"1f24,1f24,2f24" ""
"2f24,2f24,1f24" ""
"1f24,2f24,2f24" ""
"2f24,1f24,1f24" ""

"1f25,1f25,2f25" ""
"2f25,2f25,1f25" ""
"1f25,2f25,2f25" ""
"2f25,1f25,1f25" ""

"1f26,1f26,2f26" ""
"2f26,2f26,1f26" ""
"1f26,2f26,2f26" ""
"2f26,1f26,1f26" ""

"1f27,1f27,2f27" ""

"2f27,2f27,1f27" ""
"1f27,2f27,2f27" ""
"2f27,1f27,1f27" ""

"1f28,1f28,2f28" ""
"2f28,2f28,1f28" ""
"1f28,2f28,2f28" ""
"2f28,1f28,1f28" ""

"1f29,1f29,2f29" ""
"2f29,2f29,1f29" ""
"1f29,2f29,2f29" ""
"2f29,1f29,1f29" ""

"1f30,1f30,2f30" ""
"2f30,2f30,1f30" ""
"1f30,2f30,2f30" ""
"2f30,1f30,1f30" ""

"1f31,1f31,2f31" ""
"2f31,2f31,1f31" ""
"1f31,2f31,2f31" ""
"2f31,1f31,1f31" ""

"1f32,1f32,2f32" ""
"2f32,2f32,1f32" ""
"1f32,2f32,2f32" ""
"2f32,1f32,1f32" ""

"1f33,1f33,2f33" ""
"2f33,2f33,1f33" ""
"1f33,2f33,2f33" ""
"2f33,1f33,1f33" ""

"1f34,1f34,2f34" ""
"2f34,2f34,1f34" ""
"1f34,2f34,2f34" ""
"2f34,1f34,1f34" ""

"1f35,1f35,2f35" ""
"2f35,2f35,1f35" ""
"1f35,2f35,2f35" ""
"2f35,1f35,1f35" ""

"1f36,1f36,2f36" ""
"2f36,2f36,1f36" ""
"1f36,2f36,2f36" ""
"2f36,1f36,1f36" ""

"1f37,1f37,2f37" ""
"2f37,2f37,1f37" ""
"1f37,2f37,2f37" ""
"2f37,1f37,1f37" ""

"1f38,1f38,2f38" ""
"2f38,2f38,1f38" ""
"1f38,2f38,2f38" ""
"2f38,1f38,1f38" ""

"1f39,1f39,2f39" ""
"2f39,2f39,1f39" ""
"1f39,2f39,2f39" ""
"2f39,1f39,1f39" ""

"1f40,1f40,2f40" ""
"2f40,2f40,1f40" ""
"1f40,2f40,2f40" ""

```
     "2f40,1f40,1f40"    ""

     "1f41,1f41,2f41"    ""
     "2f41,2f41,1f41"    ""
     "1f41,2f41,2f41"    ""
     "2f41,1f41,1f41"    ""

     "1f42,1f42,2f42"    ""
     "2f42,2f42,1f42"    ""
     "1f42,2f42,2f42"    ""
     "2f42,1f42,1f42"    ""

     "1f43,1f43,2f43"    ""
     "2f43,2f43,1f43"    ""
     "1f43,2f43,2f43"    ""
     "2f43,1f43,1f43"    ""

     "1f44,1f44,2f44"    ""
     "2f44,2f44,1f44"    ""
     "1f44,2f44,2f44"    ""
     "2f44,1f44,1f44"    ""

     "1f45,1f45,2f45"    ""
     "2f45,2f45,1f45"    ""
     "1f45,2f45,2f45"    ""
     "2f45,1f45,1f45"    ""

1 replications
break every 45
<PermuteBalancedNoDoublets>
"Je krijgt straks steeds drie geluiden te horen.

Het is de bedoeling dat je aangeeft waar het tweede geluid
het meest op lijkt.

Als het tweede geluid het meest lijkt op het eerste geluid, klik je op 'eerste',
als het tweede geluid het meest lijkt op het derde geluid, klik je op 'derde'.

Klik om te beginnen."
"Lijkt het tweede geluid het meest op het eerste of op het derde geluid?"
"Je kunt nu een korte pauze nemen. Klik om door te gaan."
"Dit is het einde van het experiment."
0 replays
replay button 0 0 0 0 "" ""
ok button  0 0 0 0 "" ""
oops button  0 0 0 0 "" ""
responses are sounds? <no> "" "" "" "" 0    0            0
3 response categories
     0.1 0.3 0.4 0.6 "eerste" 30 "" "A"
     0.4 0.6 0.4 0.6 "tweede" 30 "" ""
     0.7 0.9 0.4 0.6 "derde" 30 "" "B"
0 goodness categories
```

7.1.5 Experiment script for the AXB task used in pilot 2

```
"ooTextFile"
"ExperimentMFC 6"
blankWhilePlaying? <no>
stimuliAreSounds? <yes>
"stimuli/" ".wav"
carrier "" ""
initial silence 0.8
inter-stimulus interval 0.4
final silence 0.8
```

```
4 stimuli
    "1f30,1f30,2f30"    ""
    "2f30,2f30,1f30"    ""
    "1f30,2f30,2f30"    ""
    "2f30,1f30,1f30"    ""

20 replications
break every 40
<PermuteBalancedNoDoublets>
"Je krijgt straks steeds drie geluiden te horen.

Het is de bedoeling dat je aangeeft waar het tweede geluid
het meest op lijkt.

Als het tweede geluid het meest lijkt op het eerste geluid, klik je op 'eerste',
als het tweede geluid het meest lijkt op het derde geluid, klik je op 'derde'.

Klik om te beginnen."
"Lijkt het tweede geluid het meest op het eerste of op het derde geluid?"
"Je kunt nu een korte pauze nemen. Klik om door te gaan."
"Dit is het einde van het experiment."
0 replays
replay button 0 0 0 0 "" ""
ok button  0 0 0 0 "" ""
oops button  0 0 0 0 "" ""
responses are sounds? <no> "" "" "" "" 0    0            0
3 response categories
    0.1 0.3 0.4 0.6 "eerste" 30 "" "A"
    0.4 0.6 0.4 0.6 "tweede" 30 "" ""
    0.7 0.9 0.4 0.6 "derde" 30 "" "B"
0 goodness categories
```

7.1.6 Experiment script for the AXB task used in the final experiment

For the practice round, this script was used:

```
"ooTextFile"
"ExperimentMFC 6"
blankWhilePlaying? <no>
stimuliAreSounds? <yes>
"stimuli/" ".wav"
carrier "" ""
initial silence 0.8
inter-stimulus interval 0.4
final silence 0.8

4 stimuli
    "a,a,u"    ""
    "u,u,a"    ""
    "i,u,u"    ""
    "i,i,u"    ""

1 replications
break every 40
<PermuteBalancedNoDoublets>
"Je krijgt straks steeds drie geluiden te horen.

Het is de bedoeling dat je aangeeft waar het tweede geluid
het meest op lijkt.

Als het tweede geluid het meest lijkt op het eerste geluid, klik je op 'eerste',
als het tweede geluid het meest lijkt op het derde geluid, klik je op 'derde'.

Klik om te beginnen."
```

```
"Lijkt het tweede geluid het meest op het eerste of op het derde geluid?"
"Je kunt nu een korte pauze nemen. Klik om door te gaan."
"Dit is het einde van het voorbeeld."
0 replays
replay button 0 0 0 0 "" ""
ok button  0 0 0 0 "" ""
oops button  0 0 0 0 "" ""
responses are sounds? <no> "" "" "" "" 0    0             0
3 response categories
    0.1 0.3 0.4 0.6 "eerste" 30 "" "A"
    0.4 0.6 0.4 0.6 "tweede" 30 "" ""
    0.7 0.9 0.4 0.6 "derde" 30 "" "B"
0 goodness categories
```

For the final experiment, this script was used:

```
"ooTextFile"
"ExperimentMFC 6"
blankWhilePlaying? <no>
stimuliAreSounds? <yes>
"stimuli/" ".wav"
carrier "" ""
initial silence 0.8
inter-stimulus interval 0.4
final silence 0.8

64 stimuli
    "1f12,1f12,2f12"    ""
    "2f12,2f12,1f12"    ""
    "1f12,2f12,2f12"    ""
    "2f12,1f12,1f12"    ""

    "1f13,1f13,2f13"    ""
    "2f13,2f13,1f13"    ""
    "1f13,2f13,2f13"    ""
    "2f13,1f13,1f13"    ""

    "1f14,1f14,2f14"    ""
    "2f14,2f14,1f14"    ""
    "1f14,2f14,2f14"    ""
    "2f14,1f14,1f14"    ""

    "1f15,1f15,2f15"    ""
    "2f15,2f15,1f15"    ""
    "1f15,2f15,2f15"    ""
    "2f15,1f15,1f15"    ""

    "1f16,1f16,2f16"    ""
    "2f16,2f16,1f16"    ""
    "1f16,2f16,2f16"    ""
    "2f16,1f16,1f16"    ""

    "1f17,1f17,2f17"    ""
    "2f17,2f17,1f17"    ""
    "1f17,2f17,2f17"    ""
    "2f17,1f17,1f17"    ""

    "1f18,1f18,2f18"    ""
    "2f18,2f18,1f18"    ""
    "1f18,2f18,2f18"    ""
    "2f18,1f18,1f18"    ""

    "1f19,1f19,2f19"    ""
    "2f19,2f19,1f19"    ""
    "1f19,2f19,2f19"    ""
    "2f19,1f19,1f19"    ""
```

"1f20,1f20,2f20" ""
"2f20,2f20,1f20" ""
"1f20,2f20,2f20" ""
"2f20,1f20,1f20" ""

"1f21,1f21,2f21" ""
"2f21,2f21,1f21" ""
"1f21,2f21,2f21" ""
"2f21,1f21,1f21" ""

"1f22,1f22,2f22" ""
"2f22,2f22,1f22" ""
"1f22,2f22,2f22" ""
"2f22,1f22,1f22" ""

"1f23,1f23,2f23" ""
"2f23,2f23,1f23" ""
"1f23,2f23,2f23" ""
"2f23,1f23,1f23" ""

"1f24,1f24,2f24" ""
"2f24,2f24,1f24" ""
"1f24,2f24,2f24" ""
"2f24,1f24,1f24" ""

"1f25,1f25,2f25" ""
"2f25,2f25,1f25" ""
"1f25,2f25,2f25" ""
"2f25,1f25,1f25" ""

"1f26,1f26,2f26" ""
"2f26,2f26,1f26" ""
"1f26,2f26,2f26" ""
"2f26,1f26,1f26" ""

"1f27,1f27,2f27" ""
"2f27,2f27,1f27" ""
"1f27,2f27,2f27" ""
"2f27,1f27,1f27" ""

2 replications
break every 50
<PermuteBalancedNoDoublets>
"Je krijgt straks steeds drie geluiden te horen.

Het is de bedoeling dat je aangeeft waar het tweede geluid
het meest op lijkt.

Als het tweede geluid het meest lijkt op het eerste geluid, klik je op 'eerste',
als het tweede geluid het meest lijkt op het derde geluid, klik je op 'derde'.

Klik om te beginnen."
"Lijkt het tweede geluid het meest op het eerste of op het derde geluid?"
"Je kunt nu een korte pauze nemen. Klik om door te gaan."
"Dit is het einde van het experiment."
0 replays
replay button 0 0 0 0 "" ""
ok button 0 0 0 0 "" ""
oops button 0 0 0 0 "" ""
responses are sounds? <no> "" "" "" "" 0 0 0
3 response categories
0.1 0.3 0.4 0.6 "eerste" 30 "" "A"
0.4 0.6 0.4 0.6 "tweede" 30 "" ""
0.7 0.9 0.4 0.6 "derde" 30 "" "B"
0 goodness categories

7.1.7 Experiment scripts for the MBEA used in pilot 2 and in the final experiment

For the first practice round, this script was used:

```
"ooTextFile"
"ExperimentMFC 4"
stimuliAreSounds? <yes>
fileNameHead = "1 Scale stimuli/"
fileNameTail = ".wav"
carrierBefore = ""
carrierAfter = ""
initialSilenceDuration = 0.5 seconds
interStimulusInterval = 0.5 seconds
numberOfDifferentStimuli = 2
  "ex1" ""
  "ex2" ""
numberOfReplicationsPerStimulus = 1
breakAfterEvery = 1
randomize = <CyclicNonRandom>
startText = "Voor dit luistertestje ga je naar melodieën luisteren.
Je hoort straks eerst twee voorbeelden, om met de testopbouw vertrouwd te raken.
Het is de bedoeling dat je pas antwoord geeft
als je beide melodieën helemaal hebt gehoord.
Klik om verder te gaan."
runText = "Zijn de melodieën hetzelfde?"
pauseText "De melodieën waren verschillend.
Klik om door te gaan."
endText = "De melodieën waren hetzelfde.
Dit is het einde van het voorbeeld."
maximumNumberOfReplays = 0
replayButton = 0 0 0 0 "" ""
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""
responsesAreSounds? <no> "" "" "" "" 0 0
numberOfDifferentResponses = 2
    0.25 0.45 0.45 0.65 "hetzelfde" "" "hetzelfde"
    0.55  0.75  0.45 0.65 "verschillend" "" "verschillend"
numberOfGoodnessCategories = 0
```

For the first testing round, this script was used:

```
"ooTextFile"
"ExperimentMFC 4"
stimuliAreSounds? <yes>
fileNameHead = "1 Scale stimuli/"
fileNameTail = ".wav"
carrierBefore = ""
carrierAfter = ""
initialSilenceDuration = 0.5 seconds
interStimulusInterval = 0
numberOfDifferentStimuli = 15
  "1" ""
  "2" ""
  "3" ""
  "4" ""
  "5" ""
  "6" ""
  "7" ""
  "8" ""
  "9" ""
  "10" ""
  "11" ""
  "12" ""
  "13" ""
```

```
  "14" ""
  "15" ""

numberOfReplicationsPerStimulus = 1
breakAfterEvery = 0
randomize = <CyclicNonRandom>
startText = "Klik om te beginnen met het eerste deel van het testje.
Het is de bedoeling dat je pas antwoord geeft
als je beide melodieën helemaal hebt gehoord."
runText = "Zijn de melodieën hetzelfde?"
pauseText ""
endText = "Dit was het eerste deel van het testje."
maximumNumberOfReplays = 0
replayButton = 0 0 0 0 "" ""
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""
responsesAreSounds? <no> "" "" "" "" 0 0
numberOfDifferentResponses = 2
    0.25 0.45 0.45 0.65 "hetzelfde" "" "hetzelfde"
    0.55  0.75  0.45 0.65 "verschillend" "" "verschillend"
numberOfGoodnessCategories = 0
```

For the second practice round, this script was used:

```
"ooTextFile"
"ExperimentMFC 4"
stimuliAreSounds? <yes>
fileNameHead = "4 Rhythm Stimuli/"
fileNameTail = ".wav"
carrierBefore = ""
carrierAfter = ""
initialSilenceDuration = 0.5 seconds
interStimulusInterval = 0
numberOfDifferentStimuli = 2
  "ex1" ""
  "ex2" ""
numberOfReplicationsPerStimulus = 1
breakAfterEvery = 1
randomize = <CyclicNonRandom>
startText = "Je hoort straks weer twee voorbeelden,
om met de testopbouw van het volgende deel vertrouwd te raken.
Het is de bedoeling dat je pas antwoord geeft
als je beide melodieën helemaal hebt gehoord.
Klik om verder te gaan."
runText = "Zijn de melodieën hetzelfde?"
pauseText "De melodieën waren hetzelfde.
Klik om door te gaan."
endText = "De melodieën waren verschillend.
Dit is het einde van het voorbeeld."
maximumNumberOfReplays = 0
replayButton = 0 0 0 0 "" ""
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""
responsesAreSounds? <no> "" "" "" "" 0 0
numberOfDifferentResponses = 2
    0.25 0.45 0.45 0.65 "hetzelfde" "" "hetzelfde"
    0.55  0.75  0.45 0.65 "verschillend" "" "verschillend"
numberOfGoodnessCategories = 0
```

For the second testing round, this script was used:

```
"ooTextFile"
"ExperimentMFC 4"
stimuliAreSounds? <yes>
```

```
fileNameHead = "4 Rhythm Stimuli/"
fileNameTail = ".wav"
carrierBefore = ""
carrierAfter = ""
initialSilenceDuration = 0.5 seconds
interStimulusInterval = 0
numberOfDifferentStimuli = 15
   "1" ""
   "2" ""
   "3" ""
   "4" ""
   "5" ""
   "6" ""
   "7" ""
   "8" ""
   "9" ""
   "10" ""
   "11" ""
   "12" ""
   "13" ""
   "14" ""
   "15" ""

numberOfReplicationsPerStimulus = 1
breakAfterEvery = 0
randomize = <CyclicNonRandom>
startText = "Klik om te beginnen met het laatste deel van het testje.
Het is de bedoeling dat je pas antwoord geeft
als je beide melodieën helemaal hebt gehoord."
runText = "Zijn de melodieën hetzelfde?"
pauseText ""
endText = "Dit was het laatste deel van het testje."
maximumNumberOfReplays = 0
replayButton = 0 0 0 0 "" ""
okButton = 0 0 0 0 "" ""
oopsButton = 0 0 0 0 "" ""
responsesAreSounds? <no> "" "" "" "" 0 0
numberOfDifferentResponses = 2
      0.25 0.45 0.45 0.65 "hetzelfde" "" "hetzelfde"
      0.55  0.75  0.45 0.65 "verschillend" "" "verschillend"
numberOfGoodnessCategories = 0
```

7.2 Vowel pairs of the target vowels

	/ɜ/		/ɐ/	
Pair number	Vowel name	Frequency (in Hz)	Vowel name	Frequency (in Hz)
1	1f1	1709	2f1	1521
2	1f2	1707	2f2	1523
3	1f3	1705	2f3	1525
4	1f4	1703	2f4	1527
5	1f5	1701	2f5	1529
6	1f6	1699	2f6	1531
7	1f7	1697	2f7	1533
8	1f8	1695	2f8	1535
9	1f9	1693	2f9	1537
10	1f10	1691	2f10	1539
11	1f11	1689	2f11	1541
12	1f12	1687	2f12	1543
13	1f13	1685	2f13	1545
14	1f14	1683	2f14	1547
15	1f15	1681	2f15	1549
16	1f16	1679	2f16	1551
17	1f17	1677	2f17	1553
18	1f18	1675	2f18	1555
19	1f19	1673	2f19	1557
20	1f20	1671	2f20	1559
21	1f21	1669	2f21	1561
22	1f22	1667	2f22	1563
23	1f23	1665	2f23	1565
24	1f24	1663	2f24	1567
25	1f25	1661	2f25	1569
26	1f26	1659	2f26	1571
27	1f27	1657	2f27	1573
28	1f28	1655	2f28	1575
29	1f29	1653	2f29	1577
30	1f30	1651	2f30	1579
31	1f31	1649	2f31	1581
32	1f32	1647	2f32	1583
33	1f33	1645	2f33	1585
34	1f34	1643	2f34	1587
35	1f35	1641	2f35	1589
36	1f36	1639	2f36	1591
37	1f37	1637	2f37	1593
38	1f38	1635	2f38	1595
39	1f39	1633	2f39	1597
40	1f40	1631	2f40	1599
41	1f41	1629	2f41	1601
42	1f42	1627	2f42	1603
43	1f43	1625	2f43	1605
44	1f44	1623	2f44	1607
45	1f45	1621	2f45	1609

This table gives the frequency of the vowel pairs, used in pilot 1, pilot 2 (pair 30) and the final experiment (pair 12-27). Vowel 1f1 and vowel 2f1 are the Slovenian vowels; these vowels were manipulated so that the frequency of the 1f-vowels decreases by 2 Hz with every pair, and the frequency of the 2f-vowels increases by 2 Hz with every pair. The F2 difference thus decreases with 4 Hz with every pair.

7.3 Duration and pitch (F0) of the vowels in the training

	Vowel	Duration in speech condition (ms)[25]	Pitch (F0) in speech condition (Hz)	Pitch (F0) in music condition (Hz)	Tone
			A1[26]	**A2**	
One	1	200	248	246.94	b
	2	230	202.5	207.65	gis
	3	200	293	293.66	d’
	4	250	253	246.94	b
	5	200	245	246.94	b
	6	230	238.5	233.08	ais
	7	200	245	246.94	b
	8	250	280	277.18	cis’
	9	200	286	277.18	cis’
	10	200	218	220	a
	11	200	250	246.94	b
	12	230	206.5	207.65	gis
	13	250	161	261.63	c’
			B1	**B2**	
Two	1	200	305.5	311.12	dis’
	2	230	286	293.66	d’
	3	200	292.5	293.66	d’
	4	230	281	277.81	cis’
	5	200	250	246.91	b
	6	250	293	293.66	d’
	7	250	356	349.23	f’
	8	200	310	311.12	dis’
	9	230	263.5	261.63	c’
	10	250	240.5	246.91	b
			C1	**C2**	
Three	1	200	252.5	246.91	b
	2	200	259	261.63	c’
	3	200	261	261.63	c’
	4	230	248	246.91	b
	5	200	228.5	233.08	ais
	6	250	268	261.63	c’
	7	200	260	261.63	c’
	8	250	234.5	233.08	ais
	9	230	236.5	233.08	ais

[25] The duration in the music condition was defined as follows: for each tone in the column ‘tone’ a duration of 250 ms was used. However, these durations were sometimes put together in the music condition, in order to make longer notes.
[26] See figure 5.

	Vowel	Duration in speech condition (ms)[27]	Pitch (F0) in speech condition (Hz)	Pitch (F0) in music condition (Hz)	Tone
	10	200	248.5	246.91	b
	11	250	251.5	246.91	b
	12	220	263	261.63	c’
	13	200	234.5	233.08	ais
	14	200	226	220	a
	15	230	310.5	311.13	dis’
	16	210	250.5	246.91	b
	17	210	221	220	a
	18	220	228.5	233.08	ais
	19	250	293.5	293.66	d’
	20	250	197.5	196	g
			D1	**D2**	
Four	1	200	252	246.91	b
	2	230	226.5	220	a
	3	200	230.5	233.08	ais
	4	200	208.5	207.65	gis
	5	250	265	261.63	c’
	6	210	287.5	293.66	d’
	7	250	236	233.08	ais
	8	220	274.5	277.18	cis’
	9	200	231	233.08	ais
	10	200	290	293.66	d’
	11	250	227	233.08	ais
	12	230	208.5	207.65	gis

7.4 Participants

7.4.1 Participants in pilot 1

Number	Sex/Gender	Age
1JV	Female	24
2TJ	Male	19
3MJ	Female	21
4JJ	Female	25
5RP	Male	22
6KG	Female	21
7RB	Female	52

[27] The duration in the music condition was defined as follows: for each tone in the column ‘tone’ a duration of 250 ms was used. However, these durations were sometimes put together in the music condition, in order to make longer notes.

7.4.2 Participants in pilot 2

Number	Languages	Sings in front of public	Instrument (years of taking lessons)	Estimation of musicality (M = musical, BM = a bit musical, NM = not musical)	Condition (1 = speech, 2 = music)	Age
1LH	Frisian (L1), Dutch (L1), English, (German)	No	Clarinet (5), Keyboard (2)	M	1	21
2MS	English, Spanish	No	Violin (7), guitar (5)	M	2	22
3BO	English, Spanish, (French) (German)	Yes	Piano (12)	M	2	26
4MO	English, Norwegian, (French), (German)	No	Piano (2)	BM	2	21
5KS	English, (German), (French)	No	Violin (8)	BM	1	26
6SC	English, Spanish	No	-	NM	1	23
7TB	English, Japanese	No	Guitar (7), drums (2)	BM	1	27
8CB	English (Russian), (French), (German)	Rarely	Hobo (7)	BM	2	20
9AS	English, German, French	No	-	NM	2	21
10JH	English, French, Italian, Spanish, German	Yes	-	M	1	31
11DB	French, English, Chinese, German	No	-	NM	1	21
12LN	English, French, German	Yes	Guitar (5)	M	1	23
13SM	English, French, Arabic (Moroccan)	Rarely	Horn (10)	M	2	21
14CJ	English, German	Yes	Clarinet (13)	M	1	23

7.4.3 Participants in the final experiment

Number	Languages	Sings in front of public	Instrument (years of taking lessons)	Estimation of musicality (M = musical, BM = a bit musical, NM = not musical)	Condition (1 = speech, 2 = music)	Age	Sex
15MW	English, German	Yes	Piano (5)	M	1	28	F
16RP	English	No	No	NM	1	21	F
17MJ	English, French, German	No	No	NM	1	23	F
18KB	English, French, German, Swedish	Yes	Flute (5), saxophone (1), guitar (0)	M	2	22	F
19KB	English, German, Dutch Sign Language (NGT)	No	-	NM	2	21	F
20MH	English, French, German	No	Piano (5)	BM	1	20	F
21YJ	English	No	-	NM	2	21	F
22BB	English, Spanish, (Swedish)	No	Guitar (0), piano (0.5)	BM	2	24	M
23SB	English, Greek	Sometimes	-	BM	1	25	F
24FK	English	No	-	NM	1	21	F
25ZK	English, (German), (French)	No	-	NM	2	26	M
26NV	English, (German), (Spanish)	No	-	M	1	26	F
27SG	English, (French), (German)	Yes	Saxophone (4), piano (0), drums (0)	M	1	23	M
28RH	English, French, (German)	Yes	Guitar (1), percussion (0)	M	2	30	M
29CB	English	No	-	NM	2	28	F

Number	Languages	Sings in front of public	Instrument (years of taking lessons)	Estimation of musicality (M = musical, BM = a bit musical, NM = not musical)	Condition (1 = speech, 2 = music)	Age	Sex
30YH	English, French, German	Yes	Piano (5), guitar (0)	M	1	21	F
31KE *not analysed*	Japanese (L1), Engels	No	Drum (0.5)	NM	1	22	F
32CS	English, French	Yes	Saxophone (9), guitar (0)	M	2	24	F
33MH	English, (German), (French)	No	Electrical guitar (0)	NM	2	23	F
34AR	English, French, Spanish	No	Trumpet (9)	BM	2	20	F
35JO	English	No	Piano (3)	BM	1	25	F
36KS	English, German, French, Polish	No	Piano (11)	M	1	29	M
37HM	English, German	Yes	Recorder (8)	M	2	21	F
38MM *not analysed*	Danish (L1), English, Swedish, Polish	No	-	NM	2	24	F
39HW	English	Yes	Piano (1,5)	M	2	23	F
40EH *not analysed*	English, German, (French)	Yes	Piano (9), bass (1,5)	M	1	21	F
41FK	English, Spanish, French, German	No	Piano (7)	BM	2	20	F
42JC	English, German	No	-	NM	1	24	F
43GR	English, German, French	No	-	NM	2	33	M
44IH	English, (French), (German)	Yes	Guitar (3)	M	1	32	M
45SS	English, (German)	No	-	NM	1	26	M

Number	Languages	Sings in front of public	Instrument (years of taking lessons)	Estimation of musicality (M = musical, BM = a bit musical, NM = not musical)	Condition (1 = speech, 2 = music)	Age	Sex
46AE	English, German, (Esperanto)	No	-	NM	2	23	M
47JB	English, French, German, (Swedish)	Yes	Guitar (0), piano (13)	M	2	24	M
48MC	English, French, German, (Mandarin)	No	-	NM	1	20	F
49IB	English, German	No	-	NM	2	23	F
50MR	English, French, German, (Russian)	Yes	-	M	1	25	F
51JE	English, (French), (German)	Yes	Violin (10), piano (2), guitar (0)	M	2	24	F
52MV	English, German, Spanish, Dutch Sign Language (NGT)	No	Guitar (7), piano (2)	M	2	21	F
53TL	English, Portuguese, (South African Dutch),	Yes	Guitar (1,5), drums (0)	M	1	26	M
54TK	English, French, (German), (Italian)	Yes	Guitar (3), piano (4), drums (1)	M	1	27	M
55FA	English, French, German	Yes	Recorder (7)	M	2	29	F
56AH	English, Spanish	Yes	Piano (3)	M	1	25	F
57LH	English	No	-	NM	1	24	M