SPEAKING IN MINOR AND MAJOR KEYS

Maartje Schreuder, Laura van Eerten & Dicky Gilbers

Running head: Speaking in Minor and Major Keys

Acknowledgements

We are grateful to Norman Cook for providing us with his cluster analysis method and to Paul Boersma for adapting PRAAT in such a way that we were able to make musical scores of speech.

Correspondence address: D.Gilbers
University of Groningen
Department of Dutch/General Linguistics
P.O.Box 716 Groningen
The Netherlands

Email: D.G.Gilbers@rug.nl
Fax: +31 50 363 6855
Tel. +31 50 363 5983

SPEAKING IN MINOR AND MAJOR KEYS

Maartje Schreuder *Laura van Eerten* *Dicky Gilbers*

ABSTRACT

In music the difference between sad and cheerful melodies is often indicated as a difference between a minor and a major key. In order to investigate whether the same difference can be found in language, we analyzed intonation contours in emotional speech. We made cluster analyses in order to find out which fundamental frequencies were most present in the contours. Furthermore, we analyzed the musical scores of sad and cheerful speech as well. In the pitch contours of all speakers we found intervals of three semitones in sad passages and intervals of four semitones in cheerful passages. We therefore conclude that emotional speech melody, just as musical melody, involves major and minor modalities.

Keywords: laboratory phonology, music, intonation

INTRODUCTION

In this paper we investigate whether or not differences in emotional speech are characterized by different modalities. Our main question is the identification of major and minor interval differences in the pitch contours of emotional speech. It is obvious that the range in the pitch contour of sad speech is much smaller than the range in

cheerful speech, but do we also speak in a minor key when we are sad and in a major key when we are happy?

Lerdahl and Jackendoff (1983), Gilbers and Schreuder (2002), Schreuder (2005) and Van Eerten (2004) a.o. observe that language and music have a lot in common. In both disciplines the research object is structured hierarchically and in each domain the important and less important constituents are defined. In that way the listener can interpret the stream of sounds. In this paper we compare intonation patterns in speech to musical melodies. One of the functions of intonation patterns and melodies is to mark boundaries. Differences in pitch movement can cause different meanings.

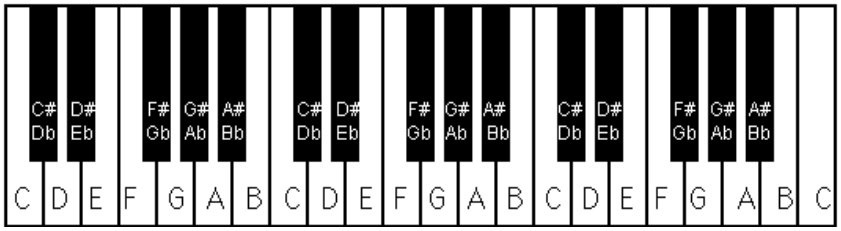
In order to investigate emotional intonation, we recorded and analyzed the performances of five professional readers reading passages from A.A. Milne's Winnie the Pooh in Dutch. In pitch contours of all speakers we found intervals between tones indicating minor modality in passages in which the sad character Eeyore is speaking and intervals indicating major modality in passages in which the happy, energetic Tigger is speaking.

This paper is organized as follows. In section 2 we outline the theoretical background; in section 3 we describe the method of our experiment and in section 4 we give the analysis and the results which are discussed in section 5.

THEORETICAL BACKGROUND

The scale in western tonal music is divided into twelve steps, also called ‘semitones’. Typical for the minor modality is that it features chords that are characterized by a distance of three semitones between the tonic and the (minor) third, whereas chords in the major modality feature a distance of four semitones between the tonic and the (major) third. This difference in thirds is the main factor for the perception of mood in music.

Figure 1 Keyboard



In Figure 1, the keys of a keyboard instrument are shown. The distance between C and C#, for instance, involves one semitone; the distance between C and D two semitones. Thus, a minor third is constituted by C and Eb and a major third by C and E. Each note has its own frequency. For example, the concert A is 440 Hz. A’ one octave higher has a double frequency: 880 Hz; A one octave lower has a frequency of 220 Hz. Within the octave, A and A’ are twelve semitones apart: five black keys and seven white keys in Figure 1. The frequency ratio between two semitones is equal. It is the twelfth root of two, which is approximately 1.0595. Table 1 shows frequency values of each note.

Table 1 Note frequencies in Hz

C	65.4 Hz	C	130.8 Hz	C	261.6 Hz
C#	69.3 Hz	C#	138.6 Hz	C#	277.2 Hz
D	73.4 Hz	D	146.8 Hz	D	293.6 Hz
D#	77.8 Hz	D#	155.6 Hz	D#	311.2 Hz
E	82.4 Hz	E	164.8 Hz	E	329.6 Hz
F	87.3 Hz	F	174.6 Hz	F	349.2 Hz
F#	92.5 Hz	F#	185.0 Hz	F#	370.0 Hz
G	98.0 Hz	G	196.0 Hz	G	392.0 Hz
G#	103.9 Hz	G#	207.7 Hz	G#	415.3 Hz
A	110.0 Hz	A	220.0 Hz	A	440.0 Hz
A#	116.6 Hz	A#	233.2 Hz	A#	466.2 Hz
B	123.5 Hz	B	247.0 Hz	B	493.9 Hz

Braun (2001) studied Dutch speech and found out that the majority of the speakers speak according to an internal tuned scale. Cook (2002) and Cook, Fujisawa and Takami (2004) investigate the modality of Japanese emotional speech. Normally, the pitch range of seven or more semitones is used in sentences. Cook et al. conclude that utterances perceived as having positive affect significantly show major-like pitch structure, whereas sentences with negative affect have a tendency to minor-like pitch structure. The conclusions are based on cluster analyses of the pitch contours of recorded utterances. In these cluster analyses the actual pitch values at every millisecond are rounded off to the value of the nearest semitone (*cf.* Table 1). The result is a graph in which one can read which semitones occur most in the utterance.

In this paper, we present a follow-up to these studies in which we try to find out whether there are different modalities in Dutch emotional speech. Cook's method has as a drawback that it is not clear whether the most frequent notes occur as direct sequences. Therefore, we will also investigate sequences of individual notes in scores of emotional speech apart from cluster analyses.

METHOD

In order to obtain different emotions in speech, we asked five primary school teachers to read selected passages in Dutch from A.A. Milne's Winnie the Pooh, in which Tigger, who is energetic and happy, and Eeyore, who is distrustful and sad, are presented as talking characters. We expect that Tigger represents a major modality, and Eeyore a minor modality in their speech.

The primary school teachers are experienced readers. The two men and three women of age 27 to 32 all claimed to have musical affinity; four of them played an instrument. They all read out the same passages, which were recorded on hard disk as wav-files and analyzed using the software CoolEdit 2000 and PRAAT (Boersma and Weenink, 1992-2005).

The passages in which Tigger and Eeyore were speaking were extracted and concatenated to ten files each varying from 8 to 53 seconds. These files were sampled every 10 milliseconds and the pitch data of each sample were obtained using PRAAT. In this way we obtained a pitch contour which we compared to the original one. Because of the great similarity, we decided that the sample rate of 10 milliseconds was sufficient for our experiment.

Subsequently, we did a cluster analysis of the pitch data in order to find out which frequencies occurred most in each contour. For this cluster analysis we relied on a model presented in Cook (2002) and Cook et al. (2004). The obtained pitch values were clustered i.e. rounded off downwards or upwards to the value of the nearest semitone. In this way, we make an abstraction of the real pitch values that can be compared to the abstractions phonologists make when they describe various allophones as the realizations of one and the same phoneme.

Furthermore, we converted the pitch contours of the stories into musical scores, to account for intervals in sequences. The aspect of time may be an important property in the analyses of modality.

ANALYSES AND RESULTS

Cluster analysis

Cook et al (2004) identifies the musical modality of Japanese speech on three peaks in the cluster analysis, because musical modality is based on triads. Nooteboom and Cohen, 1995, p.157; 162-163), however, claims that the range of Dutch intonation moves between two perceptively relevant declination levels in contrast to the three levels of English intonation. Indeed, most of our graphs show one or two peaks. There are only two graphs with three peaks. Therefore, we decided to indicate the modality on the occurrence of intervals of thirds in the graphs. If the interval between peaks concerns a minor third, we indicate the modality of speech as minor; if the interval concerns a major third, the modality is considered to be major. We define a ‘peak’ in the graph as the semitone which occurs most in a set of adjacent semitones and a ‘valley’ as the semitone which occurs least in a set of adjacent semitones.

Inspection of the cluster analyses shows that not all graphs contain more than one peak. In other words, in graphs with just one peak the modality cannot be determined. These one peak graphs were found in eight of our twenty sound files. Speech can be monotonous and neutral. Tonal music, on the other hand, has mostly a major or minor modality. In five graphs the peaks are too far apart to decide on the

modality. If the peaks constitute a fifth, for example, one cannot determine the modality.

In the seven graphs in which we do find intervals of thirds between peaks, the major modality is always observed in sound files of Tigger stories (three), whereas the minor modality is observed in sound files of Eeyore stories (four). There are no counterexamples, which indicates that our hypothesis is correct: Tigger speaks in a major key, Eeyore in a minor key.

Figure 2a shows a cluster analysis example of the raw data of Tigger as performed by subject HJ. On the x-axis the pitch values are presented in Hertz and on the y-axis the number of occurrences of a certain pitch value in the sound file is depicted. The frequency range is large, from 87 to 406 Hz.

Figure 2a Tigger in major; cluster analysis

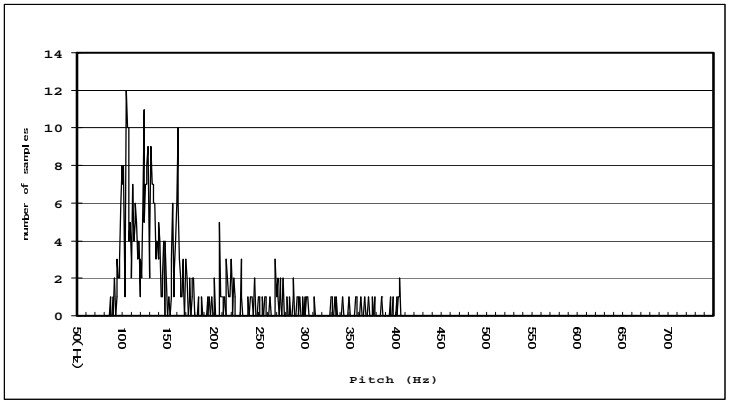


Figure 2b Tigger in major; semitones

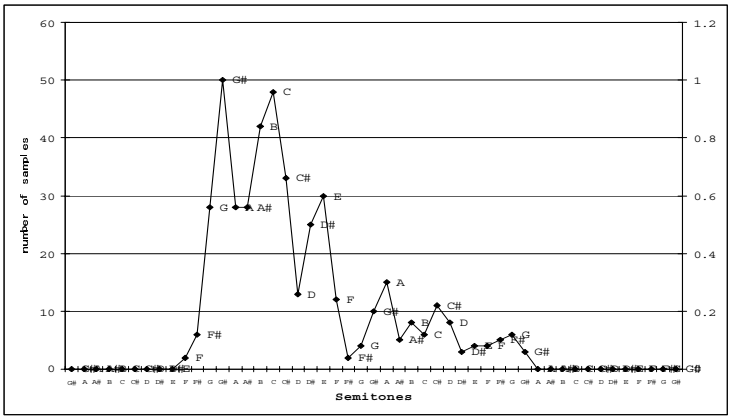


Figure 2b shows the same fragment as Figure 2a, this time clustered in semitones. The figures were obtained using Excel. On the x-axis abstractions (musical phonemes) of the real frequencies (musical allophones) are depicted as musical notes. On the y-axis we show the number of samples for each note. Our analyses are based on the semitone graphs, such as the one in Figure 2b.

Figure 2b is one of the two graphs that show three peaks. From left to right the first two peaks are on the notes G# and C. The distance of four semitones between these notes constitutes a major third. The following peak in the graph is at the note E which also constitutes a major third with the preceding C. G# and C form an inversed major third together. Tigger, as spoken by the male subject HJ, is a cheerful character and his speech indeed exhibits the major thirds of a major modality.

Figure 3a Eeyore in minor; cluster analysis

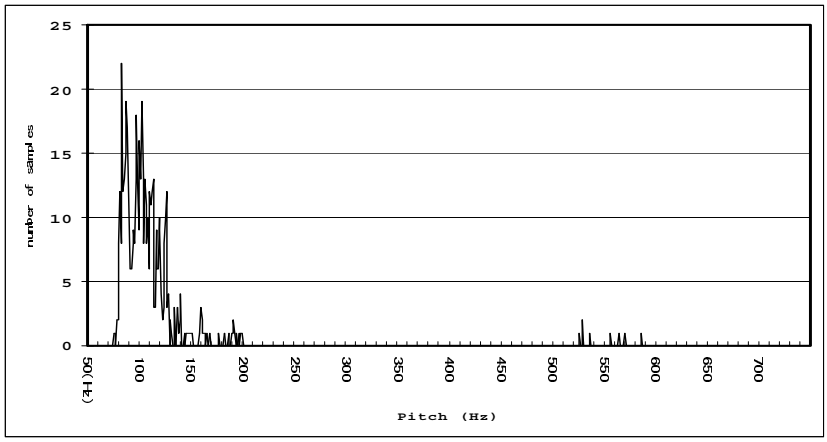
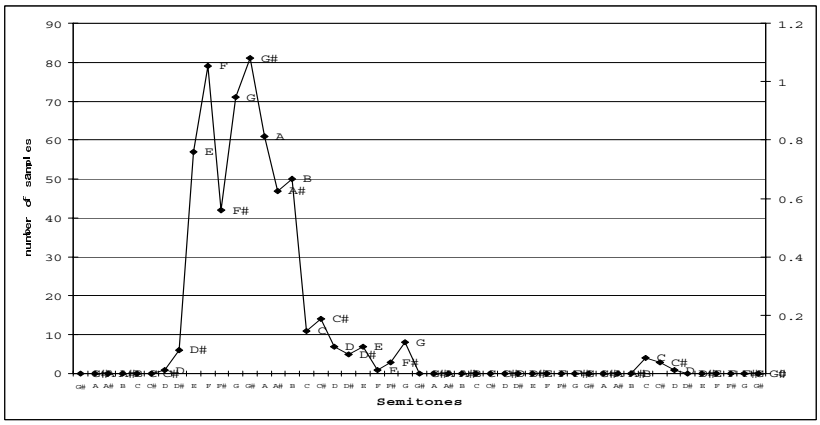


Figure 3a shows the clustered data of the same subject HJ's interpretation of Eeyore. The frequency range is smaller this time, from 75 to 200 Hz. In comparison, the frequency range of Tigger was from 87 to 406 Hz. The peaks are also located in lower regions in comparison with Tigger.

Figure 3b shows the same fragment clustered in semitones with two peaks on, respectively, F and G# (or Ab). The distance between the peaks is three semitones, in other words a minor third: Eeyore speaks in a minor modality.

Figure 3b Eeyore in minor; semitones

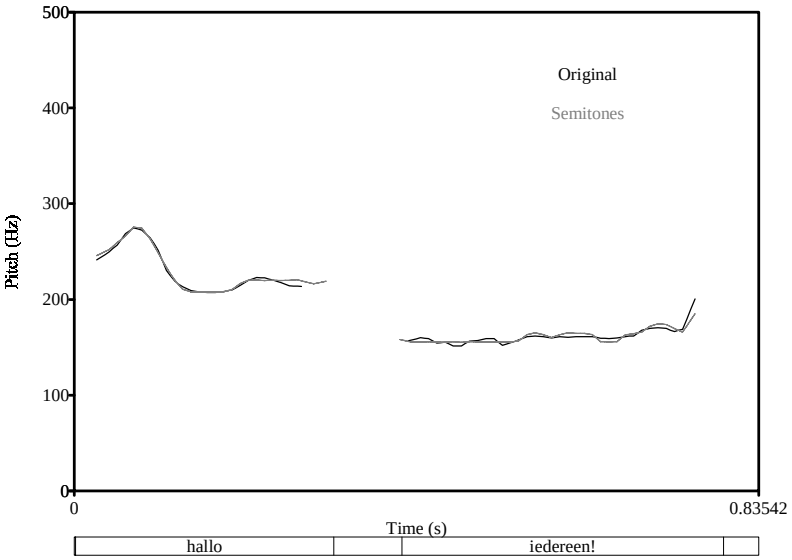


Musical scores

The cluster analysis ignores absolute intervals in time. In other words, the result is not a kind of musical score of speech. Actually, we do not know whether peaks on, for instance, C and E constitute a major third or an inversed augmented fifth. Cook (2002) justifies his choice by claiming that it is unlikely that simply an alteration in the sequence of pitches that conveys positive or negative affect could transform a minor mood into major, or vice versa (Cook, 2002, p.118). In music, however, the same melody can cause different moods depending on the chord structure of the song. For example, if a phrase in the key of C is repeated, whilst the chord progression changes to A minor, which is the parallel of C, the mood may change from cheerful to sad.

Therefore, we incorporated time as a factor, which may lead to more reliable results. We did this by using the following formula in Praat: 2 ^ (round (log2 (self/440) * 12) / 12) * 440, which works similar to a vocoder/harmonizer, rounding off automatically all frequency values at semitone value. The formula calculates the twelfth root of two for rounding off all tones to their nearest semitone, using 440 Hz, the concert A, as a reference tone. Figure 4 shows that, although this manipulation does change the original values, the differences are minor and do not reach a perceptible level.

Figure 4 Pitch contour of the original speech sound compared to the contour rounded off to the nearest semitone values (Tigger)

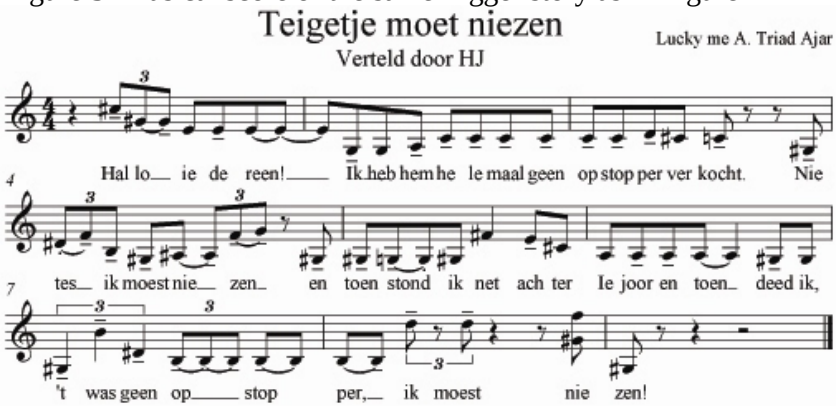


'Hello everyone!'

Subsequently, the manipulated pitch objects were resampled to sine waves. We converted these sine waves to MIDI files, using the freeware program AmazingMIDI (1998-2003). MIDI-files can be represented as musical scores by means of e.g. Steinberg Cubase software or Sibelius. In this way, the resulting musical score of a sound file enables us to determine the key and the modality of the speech.

The resulting scores of two stories, the same stories as depicted in the cluster analyses in Figures 2 and 3, are shown in Figures 5 and 6. These scores are simplified versions, because a pitch contour consists of several 'glissandos', while the MIDI-file must sample the tones into distinct notes. We chose to convert the tones into eighth notes, with the result that all notes of one glissando were unified into single chords. From these chords we chose the most prominent note for each syllable sounding in the original pitch contour. For readability reasons, the Tigger score is in the treble clef, while Eeyore spoke in a lower tone region and is therefore set in the bass clef.

Figure 5 Musical score of the same Tigger story as in Figure 2



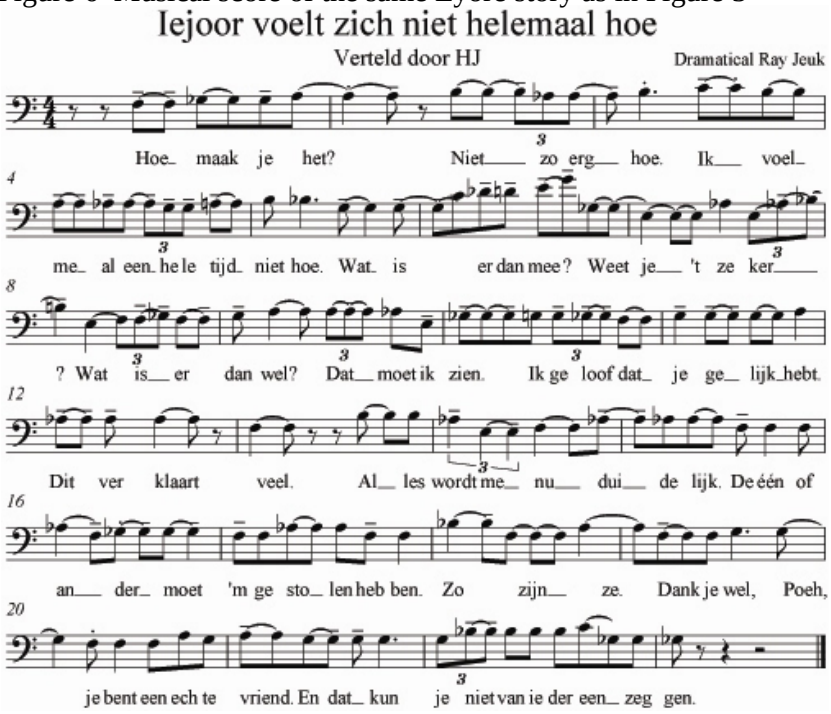
In this score of the short Tigger monologue we see the same notes stand out as in Figure 2: G#, C and E, but also A and B. A and B do not form thirds with the other notes. The objective of this score was to look whether (prominent) adjacent notes, ideally notes on neighbouring stressed syllables, form thirds in sequence. This, however, is hard to extract from the score in Figure 5, because most intervals between notes in sequences are larger intervals than thirds. Moreover, most phrases appear to be spoken on a single tone. Comparing intervals between different phrases would be wrong, because in the original speech file parts of text intervened between these phrases.

We find some thirds on stressed syllables, however, which appear to be major thirds: the interval G#-E between *lo* and *ie* in *Hallo iedereen* 'hello everyone', and the interval C#-A between *ter* and *Ie* in *achter Iejoor* 'behind Eeyore'. The major part of this score is built upon notes which form major thirds with each other. This gives the ultimate feeling of a major key: a happy, cheerful, and energetic story.

Figure 6 gives the score of the Eeyore monologue. Again we see many F's and Ab', as in the cluster analyses in Figure 3. The story is longer, and here we are able to identify sequences of thirds between stressed syllables. Examples are Gb-A in the

syllables *maak* and *het* in *hoe maak je het?* ‘how do you do?’, and F-Ab in the syllables *één* and *an* in *de één of ander* ‘one or the other (somebody)’.

Figure 6 Musical score of the same Eyore story as in Figure 3



We did not make (simplified) scores of all stories. The cluster analyses seem to give a good account of the internal relations in the melodies. While the energetic Tigger speaks in a major key, the melancholic character Eeyore expresses himself in a minor key.

CONCLUSION

In this pilot study we analyzed clustered frequency peaks in stories in which the happy Tigger and the sad Eeyore were speaking characters, and we derived musical scores of the pitch contours. The results show that in the cases in which we do find intervals of thirds between the frequency peaks, the major modality is always

observed in sound files of Tigger stories, whereas the minor modality is only observed in sound files of Eeyore stories. The derived musical scores of the intonation contours show that at least the minor thirds of Eeyore can also be found in sequences of stressed syllables.

Although speech can be neutral, a sad mood can be expressed by using intervals of three semitones, i.e. minor thirds. Cheerful speech mostly has bigger intervals than thirds, but when thirds are used, these thirds are major thirds. We conclude from this small-scale experiment that the mood of emotional prosody in speech is rather similar to musical modality. At least, the tendency we found suggests that further investigation of the similarities between music and speech could be fruitful.

ACKNOWLEDGEMENTS

We are grateful to Norman Cook for providing us with his cluster analysis method and to Paul Boersma for adapting PRAAT in such a way that we were able to make musical scores of speech. Samples of the exemplary speech and music sounds can be found on http://www.let.rug.nl/~schreudr/sounds.htm.

REFERENCES

AmazingMIDI (1998-2003). Araki Software, Japan.
<http://www.pluto.dti.ne.jp/~araki/amazingmidi/>

Boersma, P. & Weenink, D. (1992-2004). *Praat: a system for doing phonetics.*
<http://www.praat.org.>

Braun, M. (2001). Speech mirrors norm-tones: Absolute pitch as a normal but precognitive trait. *Acoustics research letters on line,* 2 (3), pp. 85-90.

Cook, N.D. (2002). *Tone of voice and mind. the connections between intonation, emotion, cognition and consciousness.* Amsterdam: John Benjamins.

Cook, N.D., Fujisawa, T. & Takami, K. (2004). Application of a psycho-acoustical model of harmony to speech prosody. *Proceedings of speech prosody.* Nara, Japan, pp. 147-150.

Eerten, L.J.A. van (2004). *Mineur en majeur in emotionele spraak, een intonatieonderzoek.* BA-thesis, Rijksuniversiteit Groningen.

Gilbers, D. & Schreuder, M. (2002). *Language and music in optimality theory*. [ROA 571-0103].

Lerdahl, F. & Jackendoff, R. (1983). *A generative theory of tonal music.* The MIT Press, Cambridge, Massachusetts, London, England.

Milne, A.A. (1994). *Winnie de Poeh.* Translation M. Bouhuys. Van Goor, Amsterdam.

Milne, A.A. (1996). *Het huis in het Poeh-hoekje*. Translation M. Bouhuys. Van Goor, Amsterdam.

Nooteboom, S.G. & Cohen, A. (1995). *Spreken en verstaan.* Van Gorcum, Assen.

Schreuder, M. (2005). *Prosodic processes in speech and music.* PhD dissertation. University of Groningen.

Schreuder, M & Gilbers, D. (2004a). The Influence of Speech Rate on Rhythm Patterns. In: Gilbers, D., Schreuder, M. & Knevel, N. (Eds.), *On the Boundaries of Phonology and Phonetics*. Rijksuniversiteit Groningen, pp. 183-202.

Schreuder, M & Gilbers, D. (2004b). Recursive Patterns in Phonological Phrases. *Proceedings of Speech Prosody 2004,* Nara, Japan, pp. 341-344.