

## Chapter 5

# Speaking in Minor and Major Keys

### 5.1. Introduction<sup>28</sup>

The prosodic phenomena discussed in the foregoing chapters were all instances of linguistic prosody. Prosody, however, also involves extra-linguistic characteristics, such as emotion. In this chapter we investigate whether or not differences in emotional speech are characterized by different modalities.

In music the difference between sad and cheerful melodies is often indicated as a difference between a minor and a major key. Our main objective is the identification of analogous interval differences in the pitch contours of emotional speech in Dutch. 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?

As we described in Chapter 1, Lerdahl and Jackendoff (1983) and Gilbers and Schreuder (2002) among others observe that intonation patterns in speech and melodies in music have a lot in common. One of the linguistic 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 (1994, 1995). We are interested in the sad character Eeyore and the happy, energetic Tigger. Although we do not find modality in the pitch contours of all speakers, we do find intervals between tones

---

<sup>28</sup> A version of this chapter also appeared as Schreuder, Van Eerten, and Gilbers (2006). The data were gathered by Laura van Eerten (2004), and she is also responsible for part of the analyses.

indicating minor modality exclusively in Eeyore passages and intervals indicating major modality exclusively in Tigger passages.

This chapter is organized as follows. In section 5.2 we outline the theoretical background; in section 5.3 we describe the method of our experiment and in section 5.4 we give the analysis and the results, which are discussed in section 5.5.

## 5.2. Theoretical Background

The difference between sad and cheerful music is often described as a difference between a minor and a major key, although in some instances composers play around with the notions of major and minor modality, which may result in cheerful music in a minor key, or sad music in the major key. The scale in western tonal music is divided into twelve steps, also called ‘semitones’. Typical of 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 57 Keyboard

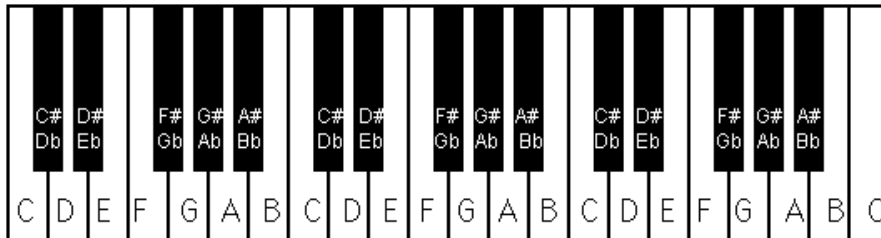


Figure 57 shows the keys of a keyboard instrument. The distance between C and C#, for instance, involves one semitone; the distance between C and D two semitones. Thus, a minor third is constituted by C and Eb and a major third by C and E.

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 57. The frequency ratio between two semitones is equal. It is the twelfth root of two, which is approximately 1.0595. Table 30 shows frequency values of each note.

Table 30 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	<b>A</b>	<b>440.0 Hz</b>
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 about seven semitones is used in sentences, a fifth. 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 down or up to the value of the nearest semitone (*cf.* Table 30).

In this chapter, we present a follow-up to these studies, in which we try to find out whether there are different modalities in Dutch emotional speech. Apart from cluster analyses we will also

investigate sequences of individual notes in scores of emotional speech.

### **5.3. Method**

In order to obtain different emotions in speech, we asked five primary school teachers to read out selected passages in Dutch from A.A. Milne's *Winnie the Pooh*, in which energetic, happy Tigger, and distrustful, sad Eeyore, are presented as talking characters.

The primary school teachers are experienced readers. The two men and three women aged 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 programs CoolEdit 2000 and PRAAT (Boersma and Weenink 1992-2006).

The passages in which Tigger and Eeyore speak were extracted and concatenated into twenty files each varying from 8 to 53 seconds. The pitch information of these files was measured every ten milliseconds using Praat. In this way we obtained sequences of frequency values representing the pitch contours. Comparison to the original pitch contours revealed a great similarity. Therefore, we decided that this sample rate of ten 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 cluster algorithm in Excel presented in Cook (2002) and Cook et al. (2004). The product of the frequency data was calculated, and assigned to the nearest semitone in an equally-tempered scale, resulting in a semitone power spectrum. In other words, the obtained pitch values were clustered i.e. rounded down or up to the value of the nearest semitone. This normalization procedure resulted in a semitone histogram in which one can read which semitones occur most in the utterance. In this way, we made 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. As Cook (2002) remarks, it might be more valid to normalize to the speaker's dominant pitches above the tonic, instead of to the musical

equally-tempered scale, and then study the interval substructure. This would probably lead to somewhat different results, but it would also complicate the analyses.

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.

## 5.4. Analyses and results

### 5.4.1. Cluster analysis

Cook et al. (2004) identify the musical modality of Japanese speech on three peaks in the cluster analysis, because musical modality is based on triads. Nooteboom and Cohen (1995: 157, 162-163), however, claim 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.

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. In contrast to tonal music, which usually has a major or minor modality, speech can be neutral.<sup>29</sup> In five cases the peaks are too far apart to decide on the modality. If the peaks constitute a fifth, for example, one cannot determine the modality. This does not immediately imply that all these instances are counterexamples, they are just indecisive. Seven cases remained for analysis.

---

<sup>29</sup> Music with a neutral modality does occur, however. Metal music, for instance, frequently uses so-called power chords, which consist solely of the tonic and the dominant. Without triads, no modality can be derived. Moreover, one can think of music without chords, with a melodic line with intervals of e.g. only fourths and fifths. This is a rare phenomenon in music, while it seems to be a normal option in speech.

Our analyses confirm our hypothesis. The major modality is exclusively found in sound files of Tigger stories in which thirds were observed, whereas the minor modality only appears in sound files of Eeyore stories. We conclude that Tigger speaks in a major key and Eeyore in a minor key in these cases.

Figure 58a shows a cluster analysis example of the raw data of Tigger as performed by subject HJ. The x-axis presents the pitch values in Hertz and the y-axis depicts the number of occurrences of a certain pitch value in the sound file. The frequency range is large, from 87 to 406 Hz.

Figure 58a Tigger in major; cluster analysis

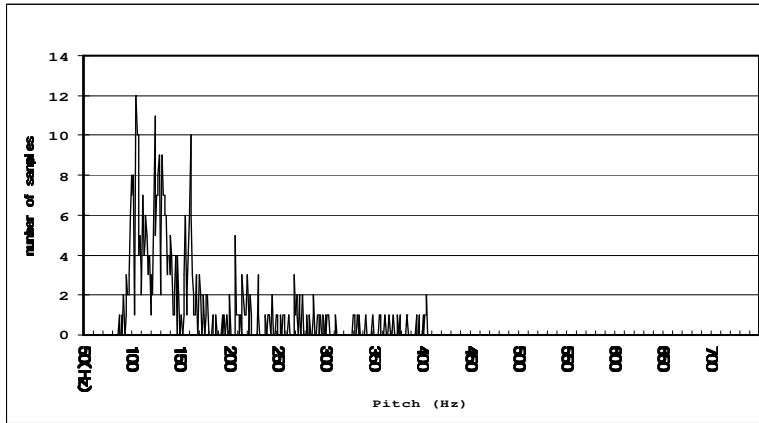


Figure 58b Tigger in major; semitones

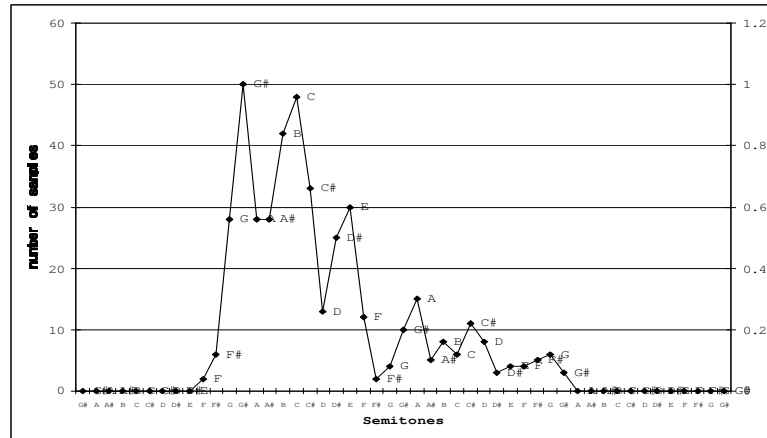


Figure 58b shows the same fragment as Figure 58a, this time clustered in semitones. The figures were obtained using Cook's cluster algorithm macro in Excel. On the x-axis abstractions (musical phonemes) of the real frequencies (musical allophones) are depicted as musical notes. On the y-axis we show the number of samples for each note. Our analyses are based on the semitone graphs, such as the one in Figure 58b.

Figure 58b is one of the few 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 inverted 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 59a 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 59a Eeyore in minor; cluster analysis

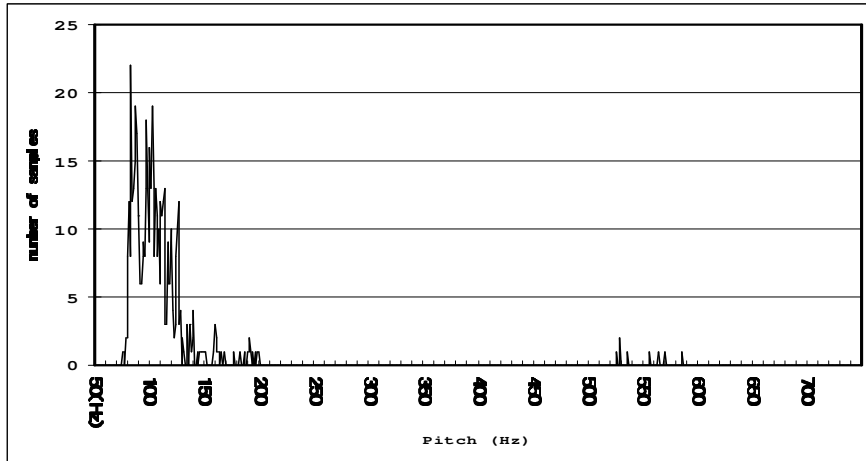
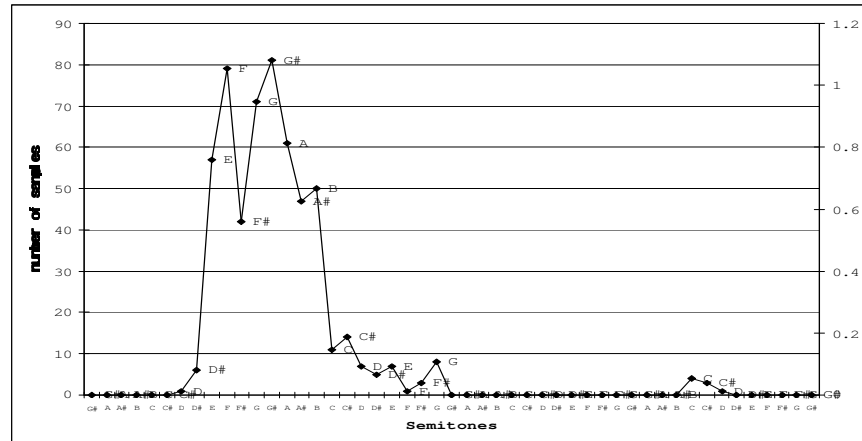


Figure 59b shows the same fragment clustered in semitones with two peaks on, respectively, F and G# (or Ab). The distance between the peaks is three semitones, in other words a minor third: Eeyore speaks in a minor modality.

Figure 59b Eeyore in minor; semitones



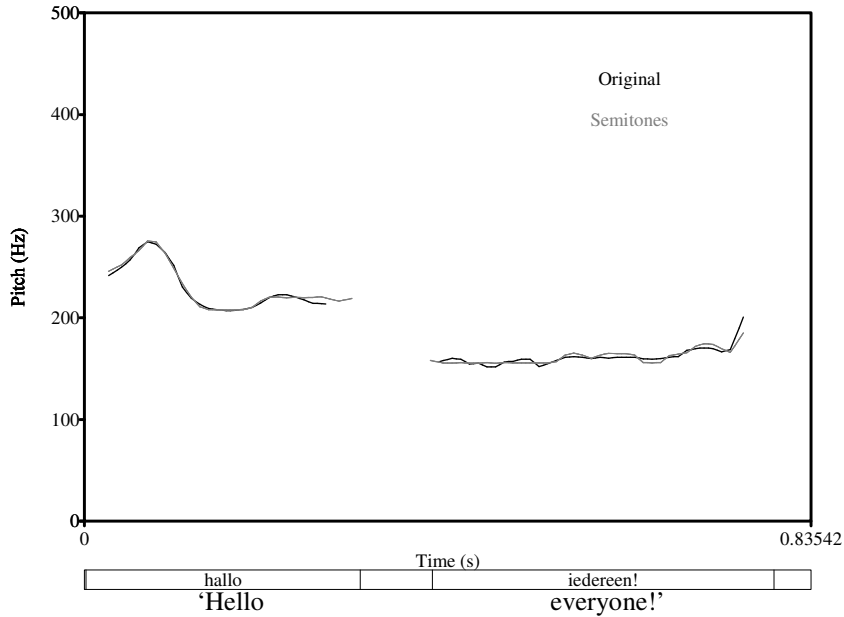
#### 5.4.2. Musical scores

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 inverted 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:  $2^{\text{round}(\log_2(\text{self}/440) * 12) / 12} * 440$ , which works similarly 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 60 shows that, although this manipulation does change the original

values, the differences are very small and do not reach a perceptible level.

Figure 60 Pitch contour of the original speech sound compared to the contour rounded off to the nearest semitone values (Tigger)



Consequently, the manipulated pitch objects were resampled to sine waves.<sup>30</sup> 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 modality of the speech.<sup>31</sup>

The resulting scores of two stories, the same stories as depicted in the cluster analyses in Figure 58 and Figure 59, are shown in Figure

<sup>30</sup> Paul Boersma made this possible by adjusting the Praat program twice to our demands. For this we are very grateful to him.

<sup>31</sup> Some MIDI-files and combined speech and musical sound files can be listened to on <http://home.planet.nl/~schre537/sounds.htm> or [www.maartjeschreuder.nl](http://www.maartjeschreuder.nl).

61 and Figure 62. 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 61 Musical score of the same Tigger story as in Figure 58

**Teigetje moet niezen** Lucky me A. Triad Ajar  
Verteld door HJ

Hal lo... ie de reen!... Ik heb hem he le maal geen op stop per ver kocht. Nie  
tes ik moest nie zen en toen stond ik net ach ter Ie jaar en toen deed ik,  
't was geen op stop per ik moest nie zen!

In this score of the short Tigger monologue we see the same notes stand out as in Figure 58: 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 61, 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 lejoor* ‘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 62 gives the score of the Eeyore monologue. Again we see many Fs and Abs, as in the cluster analyses in Figure 59. 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* ‘someone or other’.

Figure 62 Musical score of the same Eeyore story as in Figure 59

**Iejoor voelt zich niet helemaal hoe**  
Verteld door HJ Dramatical Ray Jeuk

Hoe maak je het? Niet zo erg hoe. Ik voel  
me al een hele tijd niet hoe. Wat is er dan mee? Weet je 't ze ker  
? Wat is er dan wel? Dat moet ik zien. Ik ge loof dat je ge lijk hebt.  
Dit ver klaart veel. Al les wordt me nu dui de lijk. De één of  
an der moet 'm ge sto len heb ben. Zo zijn ze. Dank je wel, Poeh,  
je bent een ech te vriend. En dat kun je niet van ie der een zeg gen.

We did not make (simplified) scores of all the 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.

### **5.5. 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 observed in sound files of Eeyore stories. Although thirds were only found in a minority of our material, there were no counterexamples in the fragments containing thirds. 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, we found a tendency that 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 tend to be major thirds. Strong conclusions cannot be drawn from only one such a small-scale experiment using a new analytic technique. But the evidence presented above is certainly suggestive. At the very least, these results are an indication that the mood of emotional prosody in speech is rather similar to musical modality. Therefore, this could be a promising method for studying emotion in speech. The tendency we found suggests that further investigation of the similarities between music and speech could be fruitful.

