Prosodic Characteristics of Emotional Speech: Measurements of Fundamental Frequency Movements

Authors: A. Paeschke, W. F. Sendlmeier

Technical University Berlin, Germany

ABSTRACT

Recent data on prosodic features of emotional speech in German are reported. Ten sentences, spoken by actors in a happy, fearful, sad, bored, angry and a neutral way served as the basis of the analyses. The features under investigation are 1. different range parameters (differences between sentence accent peak, word accent peaks, minima between accent peaks and sentence final lowest F_0), 2. the declination within phrases and 3. the duration and steepness of rising and falling F_0 movements within accents. Significant differences between emotions were found for most of the measurements.

1. INTRODUCTION

In literature the importance of prosodic features in emotional speech is evident, but has not been explored in full detail. One of the parameters which should be investigated more closely is the F_0 range. The simple measurement of range as the difference between maximum and minimum F_0 of a sentence does not convey enough information about the distribution of F_0. Patterson & Ladd [3] tested the power of four span and two level measurements with respect to their correlation with speaker states. These results suggest that the proposed range parameters may also be useful in distinguishing emotions. Similar to the pitch range model described by Patterson and Ladd, in this study the differences between maxima and minima of accents (sentence accents and word accents) and the difference between accent maxima and the final lowest F_0 value were examined for the basic emotions happiness, sadness, anger, fear and boredom.

Another characteristic feature of the F_0 movement is declination. The rising contour in some emotional utterances attracted attention, because it differs from the normal slightly falling tendency. Thus, the present study includes the analysis of declination. Here only the baseline (the line which can be imagined as the connection between the F_0 minima of a phrase) of declination was considered. The topline (the line between maxima of F_0), might also be a good parameter to differentiate between emotions and may be used in future examinations.

After having examined the steepness of rising and falling F_0 movements within syllables recently [2] now the unit of measurement is extended to accents. An accent can consist of more than one syllable and is either the sentence accent (the one which makes the strongest impression on the listener's perception) or a normal word accent. An additional examination of the final accent seems promising. In most cases the rising and falling F_0 movements cover more than the stressed syllable. Therefore, the complete F_0 movement of an accent which can expand (especially in bored utterances) up to seven or more syllables is taken into consideration. So the accent is the period between the beginning of the F_0 rise and the end of the F_0 fall around a stressed syllable.

2. SPEECH MATERIAL

Our speech material consists of high quality recordings of 10 speakers, 5 male and 5 female. All speakers taking part were professional actors. Each of them had to produce 10 sentences in a happy, fearful, sad, bored, angry and a neutral way. For the analyses in the present study three of the ten sentences were chosen:

1. Der Lappen liegt auf dem Eisschrank. (*English: The cloth is lying on the refrigerator.*)
2. Das schwarze Blatt Papier befindet sich da oben neben dem Holzstück. (*English: The black sheet of paper is up there beside the piece of wood.*)
3. Ich will das eben wegbringen und dann mit Karl was trinken gehen. (*English: I just want to take this away and then go and have a drink with Karl.*)

The sentences were semantically neutral, but it is not clear whether all sentences were equally suitable for the expression of each emotion. A maximum of three recordings per emotion and speaker were chosen for a perception test with 20 naive listeners. The listeners had to judge each recording as to the emotion they perceived. Measurements were taken only from sentences with a detection rate of at least 80%. There were 213 sentences altogether. 40 were expressing fear, 34 happiness, 34 boredom, 35 sadness, 40 expressing (hot) anger and 30 sentences were spoken in a neutral way.

3. MEASUREMENTS

All measured parameters concern specific features of pitch. These are several range measures, the declination of pitch and the duration and steepness of F_0-movements in accents.

3.1. Range

By simply measuring range as the difference between maximum and minimum of F_0 one does not get helpful information about the distribution of F_0 values within that range. Therefore, the difference between F_0 values at perceptually prominent points was determined and used to display the varieties of emotions regarding their F_0 range. To represent the F_0 range adequately the calculation of four differences seems to be sufficient. These are the differences between the F_0 maximum of the sentence accent peak and the sentence final low (1-4 in table 1), the difference between sentence accent peak and the minimum between accents (1-3), the difference between the maximum of word accent peaks and the sentence final low (2-4) and the difference between the maximum of word accent peaks and the minimum between accent peaks (2-3). For visualization of the results the first and last F_0 value of a sentence was measured. All values and differences were calculated in semitones. As a reference F_0 value for the calculation of the first and last point the lowest value of the speaking range of each speaker was taken (which can be determined from the neutral sentences). Thus, the comparability of sentences from various speakers is warranted.

3.2. Declination

Declination was first reported by Pike [5]: "The general tendency of the voice is to begin on a moderate pitch and lower the medium pitch line during the sentence". Thus, declination is defined as the falling of the maxima and minima of the F_0 curve from the beginning to the end of a phrase. A connection of the maximum F_0 values can be called the topline, a connection of the minima the baseline. In this study, the tendency of the baseline is calculated (in ST/sec.). The points used for this calculation correspond to the minima between accents of the range measurements.

3.3. Duration and steepness of accents

Three types of accents were included in this measurement: the sentence accent, the word accents and the final accent of the sentence. The sentences which express emotions characterized by a state of high arousal deserve closer attention because their final accent is remarkable.

Duration of accents: From the label files the time of the beginning and the end of an accent was extracted. This made it possible to compute the duration of all accents and the duration of the rising and falling parts of the F_0 movement over accents. For statistical interpretation a total value as well as a value for each type of accent and, additionally, the duration of the final fall of the last syllable was measured.

It is neccessary to clearly differentiate between the final accent and the final fall. The final accent can consist of more than one syllable (mostly two or three), whereas the final fall is only the falling movement in the last syllable in this experiment (the beginning of the last syllable and the beginning of the final fall do not coincide in every case).

Duration of the rising and falling parts of the F_0 movement within an accent: Due to the fact that the maximum of the F_0 within an accent lies in the middle of the accent only in a few cases, an examination of the accents with respect to the duration of their rising and falling parts will be of interest.

According to the Kiel Intonation Model accents can be characterized by early, middle or late peaks and early or late valleys. For the German language, Peters [4] reports an increased occurence of late peaks for emotional speech in general. Therefore, a mere examination of the distribution of intonation prototypes does not seem promising. The more relevant question is where exactly the peak lies. Thus, the duration from the beginning of the first syllable of an accent to the F_0 maximum of the accent was measured.

Steepness of F_0 movements: The steepness of the rising and falling parts of each accent was calculated and each type of accent (types as described above in this section) was statistically interpreted separately.

The results of the duration and steepness measurements were then viewed together in order to find characteristic features of the different emotions.

4. METHOD

The F_0 values were calculated using the ESPS/waves+ software. The necessary corrections of octave and other errors were done manually. In labeling the sentences, the type of accent and the corresponding accent peak points were determined. Sections in which the F_0 was not clearly computable (e.g. due to laryngalizations) were left out. In some cases it was difficult to determine the exact location of the start of a frequency movement; then a reasonable guess had to be made.

The F_0-values were converted from Hertz to semitones with the lower boundary of the voice range of every speaker as a reference value. The reference values were taken from the sentences spoken in a neutral way. The purpose of this normalization was to achieve results independent of the speaker. All measured parameters were statistically evaluated with an oneway ANOVA by the statistical software SPSS.

In labeling it became already evident that for some emotions there is more than one typical type as far as the progression of F_0 movements is concerned. When measuring the declination it was possible to make a distinction between the sentences with positive and negative tendency of F_0 movement. Without this distinction the positive and negative values would mix and blur the results. Within the other measurements it was not practical to create a separate examination of emotion subtypes because there would not be enough sentences remaining for each subtype to come to a meaningful statistical interpretation. Moreover, a correct classification of these subtypes is the major task for further analysis.

5. RESULTS AND DISCUSSION

5.1. Range

Table 1 summarizes the results of the measurements of the several range parameters. For each emotion the average value in semitones is shown. 1-4 is the difference between the maximum of the sentence accent peak and the sentence final lowest F_0-point. 1-3 stands for the difference between the maximum F_0 at sentence accent and the minimum between accent peaks. 2-4 is consequently the difference between a normal word accent peak and the final low. 2-3 is the difference between word accent peak and the minima between the accents.

	S	F	N	B	A	H
1-4	4,78	8,59	9,50	11,92	16,04	16,13
1-3	3,68	5,97	5,99	8,48	9,10	9,55
2-4	3,33	6,43	7,15	7,58	12,30	12,34
2-3	2,23	3,80	3,64	4,14	5,38	5,75

Table 1: *Results of the range measurements. Letters stand for the emotions: S – sadness, F – fear, N – neutral, B – boredom, A – anger, H – happiness. The numbers stand for the differences between the prominent points (for a detailed explanation see the text). If there is no vertical line between adjacent cells, the difference between them is not significant (significance level is 0.05).*

In order to illustrate these data two additional points of F_0 were determined (the starting point and the final point) and plotted in a coordinate system. The mean values of the F_0 at the beginning and at the end of sentences are listed in table 2:

	B	N	S	H	A	F
F_0b	9,97	6,72	5,97	12,64	12,52	12,38
F_0e	1,25	1,28	3,16	5,53	5,93	8,32

Table 2: *Mean F_0 values at sentence beginning (F_0b) and end (F_0e) in semitones. Reference value for the semitones calculation was the lowest F_0 value of each speaker.*

If one assumes that every sentence between the starting F_0 point and the final lowest point passes a sentence accent, several word accents and the minima between these accents with the calculated values, it is possible to draw a theoretical F_0 progression for each emotion. In such a graphic representation (see figure 1) the relations between the particular emotions regarding their distribution of F_0 and their level relative to the lower border of a speaker's voice range appears clearly.

This representation of the F_0 progression does not take into consideration the declination and the temporal specifics of the emotions because this would make the graphic too complex. Declination as well as duration parameters are discussed in detail in the next sections.

The curve representing **sadness** starts with the lowest F_0 value and has the smallest differences between maxima and minima; it does not fall considerably at the end of the sentence like most other emotions. The baseline (connection of minima) of **neutral** utterances is found around 2 semitones higher, and the baseline of **bored** utterances yet another semitone higher. Both the neutral and the boredom curves fall at their end nearly to the lowest border of the speaking range (in most sentences they fall but by the calculation of mean values the result is a little higher; for the exact value see table 2). The difference between sentence accent maxima and minima is significantly higher for bored utterances than for neutral utterances. The difference between maxima of word accents and minima is greater for boredom, but not significantly greater. Both emotions end at nearly the same point at the end of sentences, and this point is significantly smaller than in all other emotions.

Happiness and **anger** show nearly exactly the same results for all range measures but they lie significantly higher than all other emotions and have the greatest differences for sentence accents as well as for word accents. Their final F_0 value lies over the respective points for sad, neutral and bored sentences and considerably lower than the final point in fearful utterances.

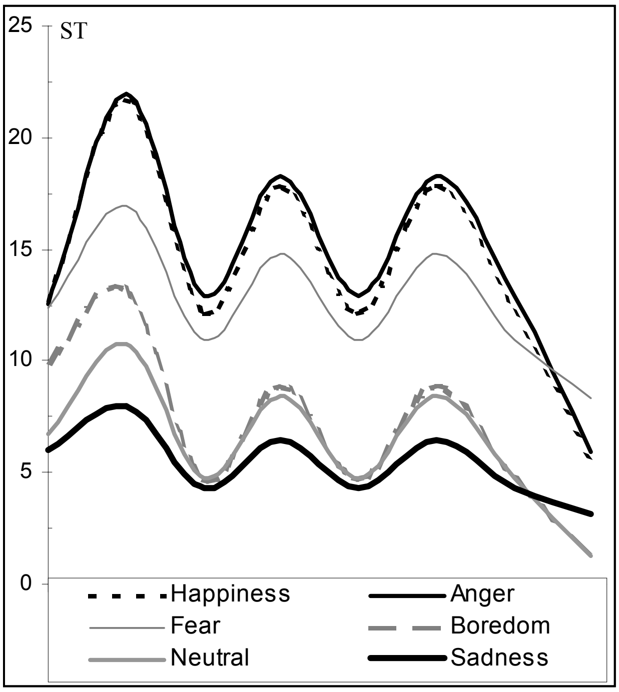

Figure 1: ***R****esults of range measurements as a theoretical progression of F_0. The illustration is a schematic model of the F_0 distribution. The number and sequence of accents (sentence accent followed by two word accents) is quite independent of their real occurrence in a single sentence.*

The differences between maxima and minima of F_0 in **fearful sentences** are almost of the same size as in neutral sentences, but the baseline of fear is twice as high as in neutral sentences. Furthermore, fearful utterances have the highest final F_0 value, which can be explained by the fact that fear relates to uncertainty. According to Peters [4] high floating F_0 movements at the end of phrases produce the impression of uncertainty, astonishment or surprise in the listener's perception. In contrast to this, the decline of F_0 to the lower border of voice range at the end of a phrase - as was found in bored and neutral utterances - is evidence of certainty, conviction and the statement of undoubted facts.

5.2. Declination

Table 3 shows the results of the declination measurement as a total value for each emotion. This procedure is useless for those emotions which occur in different versions in the database. Half of the happy sentences for example has a positive declination and the other half has a negative declination. Calculating the mean values of both types completely obliterates the facts. The declination of happy utterances does not amount to –0.88 ST/s but either to –4.67 or +4.04 ST/s (H_fall and H_rise in table 4). This also applies to angry sentences (A_fall and A_rise), whereas the distribution of falling and rising contours in angry sentences has a ratio of approximately 1:3 which causes the higher value of 1.25 ST/s in table 3.

Emotion	Number of cases	Group 1	Group 2	Group 3
Boredom	27	-4,04		
Neutral	24	-2,43	-2,43	
Happiness	23		-,88	
Sadness	22		-,65	
Anger	36			1,25
Fear	24			2,50

Table 3: *Declination of emotions in ST/S grouped according to significant differences. Negative values stand for falling F_0 tendency, positive values for rising F_0 tendency – results in different groups are significant distinct from each other (on a significance level of 0.05)*

A similar phenomenon is found for bored utterances. Although the declination is always negative, there are two siginificantly different magnitudes of declination (B I and B II). On this account a second statistical interpretation shown in table 4 seems to be more adequate. Neutral, sad and fearful utterances are rather uniform with regard to their declination; thus, a further distinction of these emotions was not necessary. **Neutral** sentences show a slight declination.

Emotion	Nr.	Group					
		1	2	3	4	5	6
B I	10	-7,26					
H_fall	13		-4,67				
A_fall	10		-3,78				
N	24			-2,43			
B II	17			-2,15			
S	22				-,65		
F	24					2,50	
A_rise	26					3,18	3,18
H_rise	10						4,04

Table 4: *Results of oneway ANOVA for declination with twofold distinction of the emotions boredom, happiness and anger grouped according to siginificant differences (values of different groups are significantly distinct from each other on a level of 0.05)*

Sad utterances have the smallest declination of all – it is hardly noticeable. The reason could be that sad utterances are accompanied by only a slight arousal and, consequently, the very low starting point of F_0 often continues over the whole sentence. Thus, a steeper declination is practically impossible.

The high value of –7.26 ST/s for bored utterances with strong falling tendency of F_0 is caused by the special performance of **boredom** in some utterances (B I). This may be characterised by the term “yawning boredom”. The actor takes a deep breath before he begins to speak and builds up a high subglottal pressure. This results in a relatively high F_0 at the beginning of the sentence and a subsequently steep falling F_0 progression up to the end of the sentence. The more frequent type of boredom in the other utterances (B II) starts with lower F_0 so that there is not enough range to perform a steep decline.

Fear is the only emotion which shows a positive declination in all sentences. The value of 2.5 ST/s is not significantly different from angry utterances with rising tendency, but it differs significantly from all other emotions.

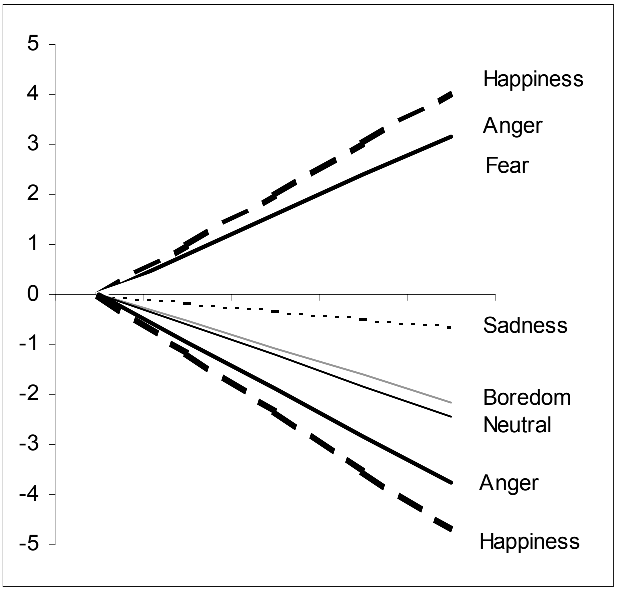

Figure 2: *Results of declination measurements. Angry and happy sentences were splitted into two groups (see the text for details).*

The results of declination measurements are similar for **anger** and **happiness** as was also the case in the results of the range measurements. The differences between them are not significant at a level of 0.05. Both emotions can be found with very steep negative as well as with very steep positive declinations. If a sentence shows an overall tendency of F_0 rise, an increasing amount of energy or an increasing power of the displayed emotion is perceived and it reaches its summit at the final accent. The opposite applies to sentences with negative declination. The greatest part of the power is observable at the first accent after which the excitation falls steadily. These results lead to the assumption that the declination is a minor point in the discrimination of the emotions happiness and anger.

5.3. Duration and steepness of accents

The parameters measured were the duration of accents, duration of the parts of accents with rising and falling F_0 movements as a total value and a separate consideration of every accent type (see also 3.3.). The same differentiation was made for the steepness measurements. All values were tested with respect to their significance with a oneway ANOVA. The used significance level was 0.05.

All measures yielded significant differences except the duration of rising parts of accents. The variation within sentences of the same emotion is probably too great, and the only way to get significant results is the use of another classification in categories relating to the shape of rise.

Duration of accents: The duration of accents in fearful utterances is significantly shorter than in all other emotional sentences, and significantly longer in bored utterances. The differences in accent duration of neutral, sad, happy and angry sentences are not significant. A separate consideration of sentence accents and word accents confirms this result. The duration of sentence accents in sad utterances is significantly shorter than in neutral ones (but still significantly longer than in fearful utterances).

Duration of falling part of F_0 movement within accents: When considering all accents collectively, the results are similar to the duration of whole accents except that the duration of final accents is of additional interest. Sad, neutral and fearful utterances have final F_0 falls of short duration (up to 300 milliseconds) and in bored, happy and angry utterances the final F_0 falls have a longer duration (from 300 ms to 1,20 sec.). These measures become more meaningful taken into account the steepness of the final accent (see below).

Duration from the beginning to the F_0 maximum of accents: When measuring from the beginning to the F_0 maximum of accents in absolute values, fearful sentences have the shortest duration (about 180 ms); a significantly longer duration is found in happy and angry sentences (230 ms) and an even longer duration in bored and sad sentences (280 ms). Calculationg the time at the F_0 maximum as a percentage value of the total accent duration, there is a significant difference between sad and bored utterances. In sad sentences the F_0 reaches its accent maximum later.

Steepness of F_0 rise: A total evaluation over all accents leads to two significantly different groups: sad, neutral and bored utterances have little steepness of rising F_0 (about 20 ST/s) - fearful, angry and happy utterances are twice as steep as the first group (about 40 ST/s). Taking into consideration only sentence accents or only word accents, the relations are the same but with higher values in sentence accents and lower values in word accents.

Steepness of F_0 fall: The F_0 falls can also be divided into the same two groups. In addition there are significant differences in the final accent and the final fall of the final accent (a description of final accent and final fall was given in 3.3.).

The falling part of the final accent in sad utterances is significantly lower than in all other emotions except bored ones. The fall of the final accent in fearful utterances is significantly greater than the fall in sad and bored sentences, but lower than in happy sentences (compare the black bars in fig. 4). The difference between fearful and angry utterances is not significant.

The results in figure 4 show a noticeable difference between the steepness of the final accent and the final fall for the emotions fear and anger. This indicates a modification of steepness within the final accent and is observable in the F_0 contour as a sharp bend.

Combining the results described above with the steepness of the final rise, the final accent in fearful, angry and happy sentences can be characterized as follows: Fearful sentences show the steepest F_0 rise in the final accent (10 ST/s higher than happiness and anger, but not significantly) whereas the fall is the lowest and shortest of these three emotions. Since the duration of the final accent in fearful sentences is extremely short, a steeper fall (like in happy and angry sentences) is barely achievable. Happy and angry sentences are both of long duration with regard to the final accent. They differ in the steepness of the final fall. Happiness is characterized by a long steady fall of about 40 ST/s. However, the final fall of angry sentences shows an increased steepness of falling on the last syllable (56 ST/s) which causes a sharp bend in the final fall.

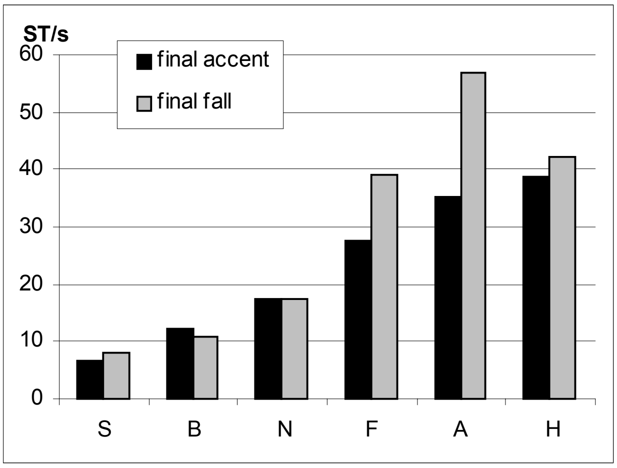

Figure 4: *Steepness of F_0 falling in the final accent and steepness of final fall (see 4.3) for each emotion in semitones per second.*

In contrast to the high values of F_0 movements of happy, angry and fearful utterances over the final accent, the movements of sad and bored sentences are rather small but nonetheless distinguishable from each other by their different length (movements in sad utterances are shorter than in bored utterances).

6. CONCLUSIONS

Generally speaking, a relatively reliable discrimination between emotions with low and high arousal can be confirmed. The results from many measurements show strong distinctions between sad, bored and neutral utterances on the one hand and fearful, happy and angry utterances on the other hand. But significant differences were also found between boredom and sadness and in particular between happiness and anger. Characterizing fear as different from all other emotions is feasible as well. Comparing for instance only the simplest parameters of range and average F_0 fearful sentences have a small range like sadness and a high average F_0 almost as high as happiness or anger. In addition, the new measurements in this study show the shortest duration of accents in fearful utterances, a significantly positive declination and the highest final F_0 value (in some sentences there is no final fall but even a slight rise on the last syllable).

The analyses in this study also revealed significant differences between sadness and boredom. Bored utterances are distinguishable by a significantly longer duration of accents than accents of all other emotions including sadness and also a longer duration of F_0 falling within the last syllable. Furthermore the F_0 differences between maximum and minimum of accents are significantly greater in bored than in sad utterances, even though the steepness of rising and falling parts of the F_0 movement within accents is not significantly different. With respect to declination, both types of bored sentences (those with slight and steep decrease) show a significantly greater downward tendency than sad sentences.

Finding parameters which allow a reliable distinction of happy and angry utterances still remains difficult. In this study, only two results of the various measurements show significant differences between happiness and anger. These are the steepness of the final fall of F_0 and the difference between maximum and minimum of the sentence accents. For angry utterances both parameters have higher values. In our future work we will continue to find more distinctive features for happiness and anger.

In this study, many different versions of sentences with regard to their F_0 movements within the same emotion were found, in particular for the sentences representing the emotions happiness, anger and boredom and to a smaller extent for fearful utterances. In order to avoid blurred results, the separate inspection of each subtype of emotion is desirable. Having found interesting results by treating subgroups of emotions separately in the declination measurements recommends a similar procedure for analyzing other parameters. The categorization of subtypes on the basis of signal parameter differences should be validated by listener's judgments.

7. REFERENCES

1. Bezooyen, R. van. *Characteristics and recognizability of vocal expressions of emotion.* Nr. 5 in Netherlands Phonetic Archives. Foris. Dordrecht. 1984.

2. Paeschke, A., Kienast, M., Sendlmeier, W. *F_0-contours in emotional speech.* Proc. of ICPhS 14:2, 929-932, San Francisco 1999

3. Patterson, D. & Ladd, R.D. *Pitch range modelling: Linguistic dimensions of variation*. Proc. of ICPhS 14:2, 1169-1172, San Francisco, 1999

4. Peters, B. *Prototypische Intonationsmuster in deutscher Lese- und Spontansprache*, In: Kohler, K.J. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung (AIPUK) Nr. 34, 1999

5. Pike, K.L., *The intonation of American English.* Ann Arbor, MI: University of Michigan Press, 1945

6. ‘t Hart, J., Collier, R., Cohen, A. *A perceptual study of intonation.* Cambridge University Press, 1990