F_0-CONTOURS IN EMOTIONAL SPEECH

A. Paeschke, M. Kienast, W.F. Sendlmeier
Technische Universität Berlin

ABSTRACT

Emotions influence a person's way of speaking, and it is possible to identify the emotional state of a speaker by merely listening to spoken utterances. The purpose of this study is to distinguish between basic emotions by prosodic features, in particular by characteristics of the fundamental frequency (F_0). In addition to the measurement of global features (e.g. average F_0 or frequency range) a number of more detailed features are examined. These parameters among others are the direction and steepness of F_0 changes on syllable level and the F_0 contour over whole utterances after a stylization procedure. Furthermore, the intrasyllabic F_0 contour without stylization is investigated. The majority of the results can be used to distinguish between the emotions fear, anger, happiness, sadness, boredom, and disgust.

1. INTRODUCTION

Numerous investigations from the past 50 years have shown that naive listeners are generally able to determine the emotional content of utterances only by hearing the speech signal. In several surveys, especially those by Scherer [7], Frick [4], Murray & Arnott [6], the acoustical correlates for specific emotions are summarized. In spite of the large amount of acoustical correlates as potential indicators for speakers' emotional states, there is little knowledge about reliable indicators for different emotions. One of the reasons is the fact that there are contradictory results from several investigations made in this area due to varying methodological strategies and a large variety of the emotions investigated.

It is widely known that emotional excitation has an important impact on speech, but the exact details of the kind of influence on the F_0 contour in particular has not definitely been explained yet. One of the few consistent outcomes of the studies mentioned above is the observation that the average F_0 increases for emotions with high excitation (such as anger, fear and happiness) and decreases for emotions with little excitation (such as sadness or boredom). Only few studies have been carried out on the emotions boredom and disgust. Hence, there is hardly any information about the F_0-contours of these emotions.

Because the syllable is one of the most important units for speech perception (see Sendlmeier [8], one may assume that also the F_0 contour of syllables could play a major role for the listener when evaluating the emotional content of spoken utterances. For this reason, an analysis on syllable basis was included to gain further insight into the structure about prosodic features of emotional speech.

2. DATABASE

Twenty sentences, spoken in six different emotional states (fear, disgust, happiness, boredom, sadness and hot anger) served as the database of this experiment. Seven actors were employed to speak every sentence with each of the six emotional contents and in a neutral way. One half of the recorded utterances consisted of single phrases, the other half of two-phrase constructions. The contents of the utterances were plain statements. Questions were left out due to their particular intonation pattern that might interact with the F_0-contour caused by the emotional content of an utterance. The recordings were rated in a perception experiment by 20 naive listeners. The task of the listeners was to determine the expressed emotion. Only the utterances with a detection rate of at least 80% were chosen for further analyses.

3. MEASUREMENTS

Initially some global parameters such as average F_0, F_0 range and standard deviation of the selected sentences were measured.

For the analysis of the F_0-contour on syllable basis the various stress levels contained in the recordings had to be considered. Four categories of stress levels - unstressed, weakly stressed, moderately stressed and heavily stressed – were differentiated (see Fant & Kruckenberg [3]). Because there is neither a practical convenient model nor an algorithm to calculate the perception of a syllable's accentuation, this step had to be done manually based on auditory impression.

According to their stress-levels the utterances were labeled and their fundamental frequency values were determined at every 10 ms using the ESPS/waves+ software and its label and F_0-functions.

Then the F_0 results were stylized according to a method developed by D'Alessandro & Mertens [2]. The stylization process provides a quantitative representation of how the prosodic attributes of a continous speech signal are perceived. As a result of this stylization only a few tonal segments per syllable remained which could be plotted as a composition of straight, rising or falling lines. These lines are assumed to reflect the auditory perception of pitch.

The syllables were categorized according to the direction of their F_0 and according to their accentuation. The distribution of the four categories of syllable stress level, the distribution of the three different F_0-contours and the distribution of different degrees of steepness for the rising and falling F_0-contours (measured in semitones per second = ST/s) were calculated.

The data were evaluated on a macroprosodic as well as on a microprosodiic level. Macroprosodic evaluation meant taking into consideration only the stylized F_0-contours whereas the microprosodic evaluation was based on the analysis of discrete F_0-values taken from the unstylized data. The investigation on a microprosodic structure is important, because the F_0 does not only influence the perception of pitch but also the perception of sound quality. Due to stylization, information about changes in sound quality is lost. Since sound quality is an important factor for the emotional classification (see Klasmeyer & Sendlmeier [5]), a microprosodic analysis was included.

Thus, the measured parameters for all neutral and emotional

utterances were: 1. global parameters such as average F_0, frequency range and standard deviation, 2. stressed syllables, 3. F_0-contours based on syllables, 4. F_0-contours of whole sentences and 5. some microprosodic features.

4. RESULTS

4.1. Global parameters of F_0

The obtained results widely confirm the results by Scherer [7], Cahn [1] and Murray & Arnott [6], indicating a connection between the emotional content and the average F_0.

Using the average F_0 in the neutral utterances as a point of reference, the results can be described as follows: Boredom and sadness were spoken with a lower average F_0 than the neutral statement, whereas disgust, anger, fear and happiness were spoken with higher average F_0 (the order of emotions given here corresponds to the increase of their average F_0).

The values of the standard deviations were the lowest for the less excited emotion sadness; a slightly higher value was found for fear and boredom. The values of the neutral recordings and the recordings representing disgust have medial positions with values of 2,6 semitones (ST). The highest values were obtained for the angry and happy utterances. This confirms our expectations that higher standard deviation indicates a greater fluctuation of the F_0, also expressed in the larger number of direction changes of the F_0-contour found in the utterances representing happiness and anger.

4.2. Stressed syllables or accentuation

Looking at the distribution of the differently stressed syllables, specific features were found for the different emotional states. The distribution of the four degrees of syllable stress level in **neutral** utterances is as follows: Nearly two thirds of the syllables are unstressed, one third is weakly stressed and the rest are either moderately or heavily stressed. The distribution for **bored** utterances is exactly the same. In **sad** utterances the distribution of stressed syllables is similar to sentences which are spoken in a neutral or bored way. They have the same number of unstressed syllables, but have more weakly and fewer heavily stressed syllables. All other emotions (happiness, anger, fear and disgust) have considerably fewer unstressed syllables. **Disgust** has the largest number of weakly stressed syllables. **Anger** has the lowest rate of unstressed and the most heavily stressed syllables. **Fear** and **happiness** have similar distributions of stressed syllables: One third is unstressed, one third is weakly stressed and another third is moderately/heavily stressed. But they are different in the proportion of the moderately and heavily stressed syllables such that fear has equivalent numbers of moderately and heavily stressed syllables, whereas happiness has only a few heavily but many moderately stressed syllables.

4.3. Syllable based F_0-contours

4.3.1. Direction of F_0-contour. The emotional utterances were analyzed with respect to the directions of their F_0-contours. The stylization showed that almost all syllables consisted of only one tonal segment,i.e. either a) a constant segment, b) a rising segment or c) a falling segment.

The distribution of the different segments for the recorded utterances can be used to distinguish between their emotional contents. The largest amount of syllables with constant F_0 is found for the utterances representing boredom, sadness and fear (about 75%). That is about 5% more than the amount of syllables with constant F_0 in neutral syllables. Disgust, happiness and anger have about 50% syllables with level F_0.

The amount of falls and rises in neutral utterances is relatively small (13% each). All emotions except fear have a larger amount of syllables with falling F_0-contours. Thus, the small amount of syllables of such a falling contour can be used as an indicator of the emotion fear. In contrast to this, characteristic for sadness and boredom is a very small number of syllables with rising F_0-contours. The emotional states disgust, happiness and anger can be characterized by high numbers of occurrences of rising and falling F_0-contours. Many rises but some fewer falls are typical for happiness.

4.3.2. Steepness of rising and falling F_0. Another interesting point is the steepness of rising and falling of the fundamental frequencies of the syllables. Generally speaking the quotient of ascending and descending F_0 is greater than 1, regarding all emotions except for boredom where the quotient is between 0 and 1. The least steepness is found for boredom, the second least steepness for sadness (both smaller than 10 ST/s). Anger reached the highest values for the steepness of rises and falls (>30 ST/s). The steepness of rises in neutral sentences, utterances expressing fear, disgust and happiness is around 25 ST/s. The steepness of falls is somewhat lower, i.e. about 20 ST/s in neutral utterances and utterances expressing disgust. Sentences which were spoken happily reached more than 20 ST/s and utterances spoken with fear only 14 ST/s.

4.3.3. A detailed view on progression of F_0 with respect to accentuation. So far the analysis of the direction and steepness of the F_0 for all four degrees of accentuation were carried out together. When looking at stressed and unstressed syllables separately further interesting results were found.

4.3.3.1. Syllables with rising F_0. In emotional sentences the steepness of rising syllable contours for stressed syllables is greater than for unstressed syllables. In neutrally spoken sentences this is reversed. The results can be summarized as follows: For **fear, happiness and disgust** the steepness in stressed syllables is about 30 ST/s, in unstressed syllables about 15 ST/s. For **anger** the steepness in stressed syllables is about 30 ST/s, in unstressed syllables about 25 ST/s. There is a directly proportional relation between the steepness of the rise of the F_0 and the degree of accentuation for fear and anger. The emotions happiness and disgust are special, because their greatest steepness is reached in moderately stressed syllables and not in heavily stressed ones. **Boredom** and **sadness** have a hardly noticeable steepness when syllables are unstressed and a little steepness when they are stressed.

4.3.3.2. Syllables with falling F_0. It is remarkable that in neutral utterances and in the sentences expressing fear and disgust, moderately and heavily stressed syllables were not spoken with falling F_0. **Anger**, **fear** and **sadness** show the same steepness of falls for all four degrees of syllable stress. Sadness has little steepness (<10 ST/s), fear has moderate steepness (>15 ST/s) and anger has great steepness (>25 ST/s). For the other emotions we found that the stronger the stress, the smaller is the steepness of falls. This applies to **boredom** which shows little

steepness of falls, to **happiness** which shows great steepness of falls, and to **disgust** which shows great steepness of falls in unstressed syllables and little steepness of falls in weakly stressed syllables (moderately and heavily stressed syllables with falling F_0 do not exist).

4.4. F_0-contours of sentences

Typical contours of emotional utterances are plotted in Figure 1. The picture refers to the second phrase of the utterance "Die wird auf dem Platz sein, wo wir sie immer hinlegen" ("It will be in the place where we always put it."). For the utterance expressing disgust a different phrase was chosen ("Das Brett habe ich angebunden" - "I have tied the board up."), due to the fact that too few listeners recognized disgust when listening to the other sentence. The thick lines indicate the stylized F_0 of the emotional utterances, and the thin lines represent the corresponding neutral version of the utterances. The main accent of the sentence lies on the first syllable of the word "immer" (always). Although the first syllable is stressed here, the maximum F_0 is reached on the second syllable. This applies to all emotions except anger.

Anger: The progression of F_0 differs from the neutral expression in three ways: It progresses on a higher level, its rises have a greater steepness, and the maximum F_0 is reached later than that in the neutral utterance. Furthermore, the time the main stress takes (from the beginning of the rise up to the maximum of the F_0 and further to the end of the fall) is shorter than in neutral utterances. (In the figure this does not show clearly because the syllables are time synchronized).

Fear: When examining fear, a higher level of the F_0 (still higher than in anger) and a greater steepness of rises was found. But the most characteristic feature of fear is that the F_0 remains on this high level up to the end of the sentence, whereas all other sentences show the typical final fall.

Happiness: In happy utterances more syllables are included in the main accent of the sentence. That means that the rise of the F_0 begins earlier time and forms a wide rounded summit. In addition, the syllables are often composed of several segments and form a contrast to neutral utterances which mostly have syllables with only one segment. A considerably increased level of the F_0 is remarkable, too.

Boredom & Sadness: These two emotions have very similar F_0-contours. Both are spoken with the fewest changes of the F_0. Irrespective of some laryngalizations and some light emphases the progression of the F_0 almost looks like a sequence of equal pitch levels. The only major difference between these emotions is the duration of the syllables. Sad utterances have only a few considerably lenghtened syllables and many syllables of normal duration. Bored utterances mostly have moderately lengthened syllables and only a few normal ones.

Disgust: Basically the F_0-contour is similar to that in neutral utterances, but on a slightly lower level with larger discrepancies (e.g. the lower level in "das" and the higher level in "ge") and even abnormalities every now and then. In the example, there are many laryngalizations which cause the unnatural rises in the syllables "ich", "an" and "den". In addition, in disgust the duration of the syllables is always longer than that of the neutral syllables (note the different time scale in this part of the picture).

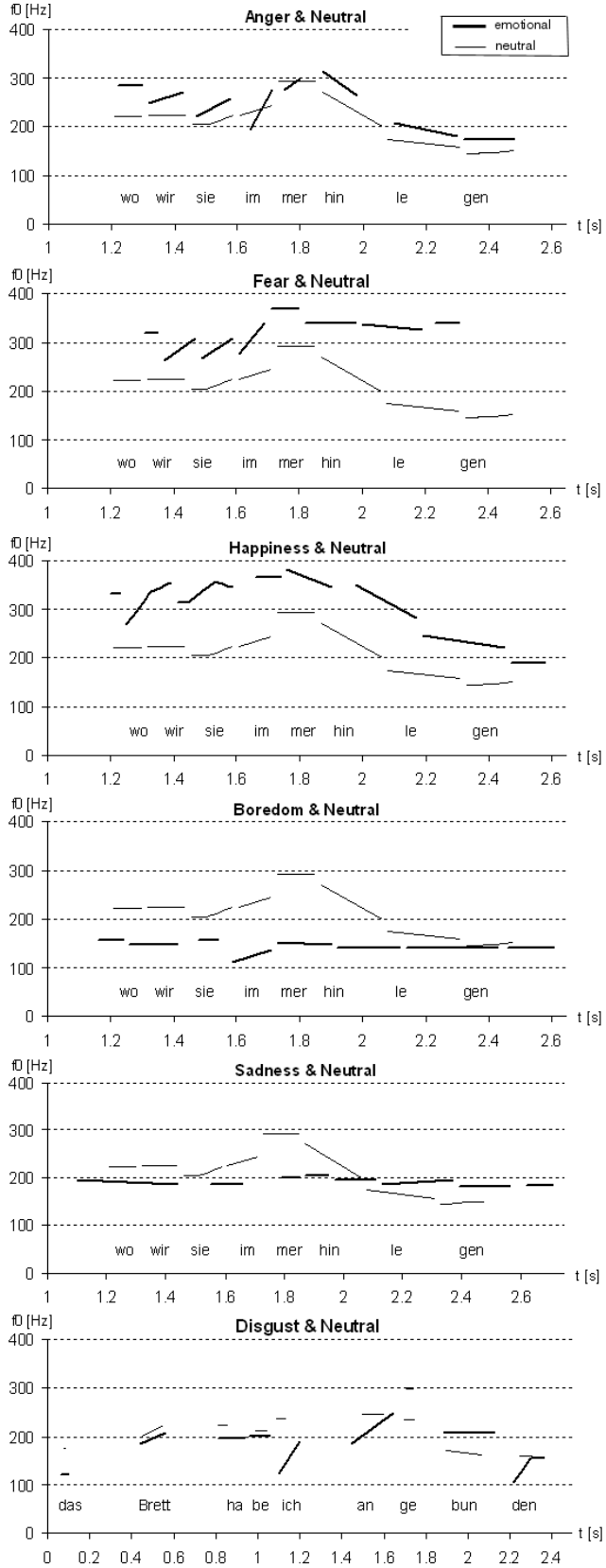

Figure 1. Stylized F_0-contours of emotional and neutral utterances

4.5. Special microprosodic features

In normal voice the progression of F_0 on one syllable is continuous, a nearly straight line (upward or downward in the case of a rising or falling frequency) with only little deviations. But in the case of emotional excitation several kinds of perturbation can occur. Figure 2 shows two special appearances in happy and sad utterances which sound nearly like laughter and howl. The progression in the happy syllable is similar to a sine wave, whereas the sad syllable seems to be more irregular.

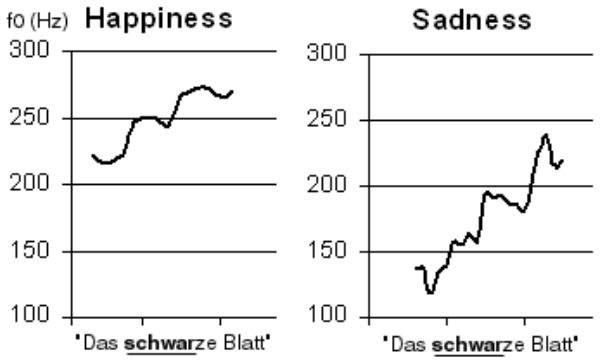

Figure 3 shows the influence of laryngalization on the following progression of the F_0. Clearly noticeable is the abrupt change between laryngalization and normal frequency for the neutral and happy examples. In syllables of angry utterances there is a short rise from the lowest to the normal point followed by the normal progression of slightly falling F_0 (it is the last syllable of the utterance). But the sentence spoken with disgust has a continous transition from the laryngalized low frequency to the normal one. After attaining the maximum frequency, it immediately falls to another laryngalized low frequency.

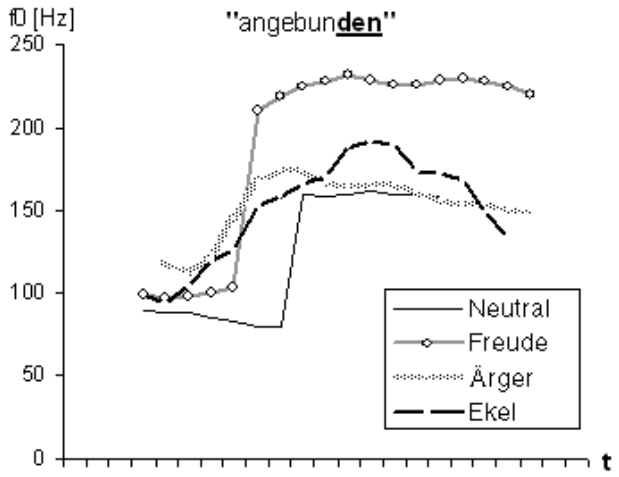

Figure 3. Influence of laryngalization

5. DISCUSSION

The results of this investigation show a number of specialties concerning the different emotions. Each measured item reveals particular characteristics of the emotions and makes it possible to distinguish between them.

In summary the results are: 1. The average F_0 of the six emotions investigated differ from each other. 2. Also, their degree of stress differs. The intensity of syllable stress was found to serve as an indicator for the different emotional states of the speaker, e.g. in emotions with little excitation like sadness the number of strongly accented syllables is reduced in contrast to emotions with high excitation like anger. The number of unstressed, weakly, moderately and heavily stressed syllables contained in an utterance results in a specific pattern for each emotion. This allows for the assumption that emotions could be determined by their own rhythm. 3. The direction and steepness of the F_0 of syllables were also found to be emotionally specific. This needs to be confirmed by further measurements including more sentences and more speakers. The clearest result is the unusually small number of syllables with falling F_0 in fearful utterances in comparison to neutral and other emotional utterances. 4. Conspicuous effects were obtained for the F_0-contour throughout whole phrases and sentences. In addition, special features of non-stylized progressions of the F_0 were found in utterances expressing happiness, sadness and disgust.

Future work should concentrate on measuring further features of patterns of the F_0, especially with regard to the contours of sentences and a number of other prosodic characteristics. These will include the investigation of synchronous or asynchronous appearance of the maximum F_0 and the maximum energy in syllables and other appropriate parameters to determine characteristics of duration, rhythm and loudness. The final aim is to develop emotional profiles consisting of as few as possible prosodic parameters, with the help of which one may differentiate between the basic emotions anger, fear, happiness, sadness, boredom and disgust.

ACKNOWLEDGMENT

This work was supported by the Deutsche Forschungsgemeinschaft DFG (German Research Community), grant nr. SE 462/3-1 to W.F. Sendlmeier.

REFERENCES

[1] Cahn, J.E. 1990. Generating expression in synthesized speech. *Technical Report, MIT, Media Technology Laboratory*, MA, USA

[2] D'Alessandro, C., Mertens, P. 1995. Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language*, 9, p. 257-288

[3] Fant, G. & Kruckenberg, A. 1989. Preliminaries to the study of swedish prose reading and reading style. *STL-QPSR* 2, 1-69

[4] Frick, R.W. 1985. Communicating emotions: The role of prosodic features. *Psychological Bulletin* 97, p. 412-429

[5] Klasmeyer, G. & Sendlmeier, W. 1999. Voice and Emotional States. In: R. Kent & M. Ball (eds.) The Handbook of Voice Quality Measurement. *Singular Publishing Group*, San Diego, in press.

[6] Murray, I.R., Arnott, J.L. 1992. Towards a simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *JASA* 93 (2), p. 1097-1108

[7] Scherer, K. 1986. Vocal affect expression: A Review and a Model for Future Research. *Psychological Bulletin* 99, 2, p. 143-165

[8] Sendlmeier, W.F. 1995. Feature, phoneme, syllable or word: How is speech mentally represented? *Phonetica* 52, 131-143.