Global Trend of Fundamental Frequency in Emotional Speech

Astrid Paeschke

Department of Communication Science
Technical University of Berlin, Germany

paeschke@kgw.tu-berlin.de

Abstract

In this study – which is part of an extensive investigation of the prosodic features of emotional speech – global trends of fundamental frequency were examined. The primary goal was to test the use of this parameter to characterize a specific set of emotions (happiness, anger, anxiety, sadness, disgust and boredom). Global F_0 trend was measured in the form of the gradient of the linear regression in order to avoid the disadvantages associated with the determination of base and top lines commonly used to examine declination. A secondary goal was to bring a new argument into the discussion of causes for declination and its degree of influence on speech production.

Results show a significant steeper downward trend for boredom than for neutral speech and a significant smaller falling global trend for all other emotions than for neutral speech. While the global trend seems to be appropriate especially for the description of boredom and sadness and to a slightly lesser extent also for anxiety and disgust, it is almost meaningless for the emotions happiness and anger. In happy and angry utterances the pitch accents are strong enough to hide the global downward trend completely. The results for each emotion are discussed in detail with respect to specific physiological constraints, as well as auditory impression and other prosodic features typical for these emotions.

1. Introduction

During the last decades interest in emotions and their acoustical correlates has steadily increased. Even though fundamental frequency has often been measured (for reviews see [7], [13]) the declination of emotional utterances has never been examined. An investigation of general trends of fundamental frequency over the course of utterances could not only be useful in the field of emotion research but can also contribute to the understanding of declination. Because of the peculiar physiological conditions of emotional states, the analysis of emotional speech is a promising approach for the examination of the declination phenomenon. Results of a study by Swerts et al. ([25]), who found differences in declination for different speaking styles, support this assumption.

1.1. Declination

The term *declination* was first used by Pike ([19]). He observed a general downtrend of fundamental frequency over the course of an utterance. His observation was confirmed in a number of following studies but the causes of declination, its relevance as a carrier of linguistic, para- and extralinguistic information and even its existence are still controversial. Only the fact of a gradual fall of the subglottal pressure is generally accepted as a main factor for the declining fundamental frequency over the course of an utterance.

A number of theories – differing in details – hypothesize an automatic declination mechanism inherent to speech production (e.g. [2], [4], [5], [10], [11], [30]). Other researchers argue that there is no such mechanism – declination is only the result of intonational variation. Decreasing fundamental frequency movements are merely caused by "downstepping" (see Pierrehumbert [19]). For a detailed discussion of the literature about declination see Ladd [9]. Ohala (p. 42 in [14]) remarks "that the question of whether F_0 declination is caused by laryngeal or by respiratory activity has still not been answered definitively."

Tatham and Lewis ([27]) suggest a combination of these two opposed views. They assume the existence of a declination based on the changes of subglottal pressure which can be influenced by the speaker. They define declination as an intrinsic process which can be systematically modified by the speaker to produce inclination (over more than a single word) as well as declination of the fundamental frequency. They claim that speakers are able to exert enough influence on the subglottal pressure to push it up or down.

Morton et al. ([12]) note that there is still demand for a suitable representation of global movements of fundamental frequency. The task for further experiments will be to find out which parts of specific fundamental frequency movements are caused by declination and which are caused by a deliberate composition of the intonation contour by the speaker.

1.1.1. Determination of declination lines

Apart from studies where the declination of the fundamental frequency is just determined by viewing the graphical representation of the fundamental frequency curve (e.g. [29]) the simplest way to determine declination lines (usually base and top lines) is to draw a line connecting the minima and maxima of the fundamental frequency curve (e.g. [2], [3], [11]). There are substantial objections to this procedure. First of all, how can one draw a line if the peaks (or valleys) do not fit into a straight line? This situation is rather the norm than the exception. Another problem occurs when there is only one peak in the phrase – for a top line at least two points are needed.

Further disagreement prevails with respect to the time-dependence of the declination. Is the degree of declination dependent of the length of an utterance or not? This also includes the question if the level of the fundamental frequency at the beginning of an utterance is concerned. Some ambitious approaches therefore use logarithmic scaled data ([28]) or allow a sharp bend ([6]) in the declination line to map a changing declination from a steeper decrease at the beginning of an utterance to a lesser decrease of the fundamental frequency later in the utterance, even though this does not solve the basic problems.

1.2. Terms: Global Trend vs. Declination

The term "declination" implies a decrease of the fundamental frequency over the course of an utterance. Though this is appropriate for neutral speech, emotional speech is different. In emotions characterized by high arousal, extreme pitch accents can be observed which overlay or disable the general downtrend – in particular if they are located at the end of an utterance. These pitch accents cause an increase in fundamental frequency over the course of an utterance. That is one of the reasons for the use of a different term in this study. The term "global trend" can describe a movement of fundamental frequency in any direction.

Another reason for avoiding the term "declination" is the inconsistency in definition of declination and in methods for determination of declination lines. According to different theoretical assumptions about the causes for declination, the term is used either for a phenomenon inseparable from speech production and incontrollable by the speaker, or for the conscious setting of pitch accents actively controlled by the speaker. Neither assumption eliminates the possibility of both factors acting together.

Given such discrepancies, global trend of fundamental frequency, as suggested here, shall be defined as the gradient of the linear regression, calculated as an objective measure from the statistical procedure of simple linear regression. In the interpretation of the results, note that this kind of procedure does not allow us to tell apart the influences of speech mechanisms and intended intonation. However, the results, considered together with knowledge of physiological characteristics of emotions and other intonation features of emotional utterances, make it possible to draw conclusions about the degree of influence of physiological factors of speech production on the course of the fundamental frequency.

2. Speech Material

A database of more than 500 utterances served for the analysis. The utterances were spoken by actors who performed 10 different sentences in the emotions happiness, anger, sadness, anxiety, disgust and boredom as well as in a neutral version. All utterances were evaluated regarding their emotional content and naturalness in a perception experiment with 20 listeners (for more details see [15], [16] or [17]).

3. Measurements

In the auditory analysis, different shapes of movements of the fundamental frequency for the different emotions were clearly observable. Therefore the attempt of exploring and describing these impressions by acoustical measurements seemed worthwhile.

Global Trend was measured using the linear regression method. If there was a reset observable in the frequency excursion of an utterance, the regression line was calculated for each part separately. The measurement was taken automatically by a Matlab script. In order to represent human perception as well as possible, calculated values were converted from Hertz to semitones.

Additionally the time-dependency of the global trend was determined by calculating the correlation between utterance duration and global trend. Results are shown in table 1.

4. Results and discussion

Table 1 shows the calculated values of the global trend for every emotion and for neutral speech. The statistical analysis (one-way ANOVA) showed significant differences between neutral speech and all emotions. Neutral speech is regarded here as a reference value and shows a global downtrend of about -3 semitones per second. In bored utterances a stronger downfall than in all other speaking styles was found (-4 semitones per second). The mean global trend for happy utterances is nearly -2 semitones per second. All other emotions (anger, anxiety, sadness and disgust) feature only small global trends of around -1 semitone per second.

Table 1: *Global trend of fundamental frequency in semitones per second. Hyphens between rows indicate significant differences between emotions regarding global trend (p=0.01). Values in the third column specify the correlation coefficient for the correlation between global trend and utterance duration.*

Emotion	Global trend of F_0 (in ST/s)	Correlation r
Boredom	-4,01	0,79
Neutral	-2,93	0,66
Happiness	-1,89	0,42
Disgust	-1,19	0,57
Anger	-1,19	0,46
Anxiety	-0,81	0,40
Sadness	-0,84	0,51

But these values show only half of the truth. One can get much deeper insight into real circumstances by looking at the distribution of the values. Figure 1 shows the distribution of the trend lines in the form of histograms. In the following the characteristics of each emotion are explained in detail.

Anger: The observed variation in global trend values was the greatest for anger. The distribution ranges from +7 to –8 semitones per second and resembles the Gaussian distribution adequately. A possible explanation for the wide distribution of global trend values, and especially the appearance of upward trends, lies in the specific speaker's strategy for expressing anger: The speaker is loaded with anger and got exact one sentence to let out all this piled up energy. The database contains three types of disruptions. They differ in the location of the outburst (beginning, middle or end of the utterance). This results in one or more extremely high pitch accents at the respective position and accordingly to a falling (peak at the beginning), straight (peak in the middle) or upward global trend (peak at the end of the utterance).

Happiness: Trend lines for happiness are between –9 and +3 semitones per second. For this emotion the same applies as for anger. It seems that there is not just one prototypical trend, but the speaker does have enough energy to configure the progress of fundamental frequency to suit himself. The many fundamental frequency rises are big enough to overlay the general declination connected with normal speech production.

Sadness: The results for sadness show the smallest variation of global trend values at all. Nearly all utterances show a

straightforward trend. Considering the fact that sadness has a very narrow fundamental frequency range the global trend line already represents the effective progression of the fundamental frequency through the utterance very well. Low arousal, and the minimal tension of the larynx muscles associated with sadness, are plausible causes for this. Because of the lower vocal effort, air consumption is small, and the subglottal pressure (which is seen as the major cause for declination of fundamental frequency) decreases more slowly than in neutral speech. As a result the gradient of the regression line is very small. In addition, a larger drop in fundamental frequency would not be possible, because speakers start their utterances little higher than the lowest fundamental frequency limited by their speech organs. Therefore they often reach the lowest fundamental frequency before the end of an utterance and continue speaking with creaky voice. Sometimes they increase F_0 a few Hertz after reaching the lowest point.

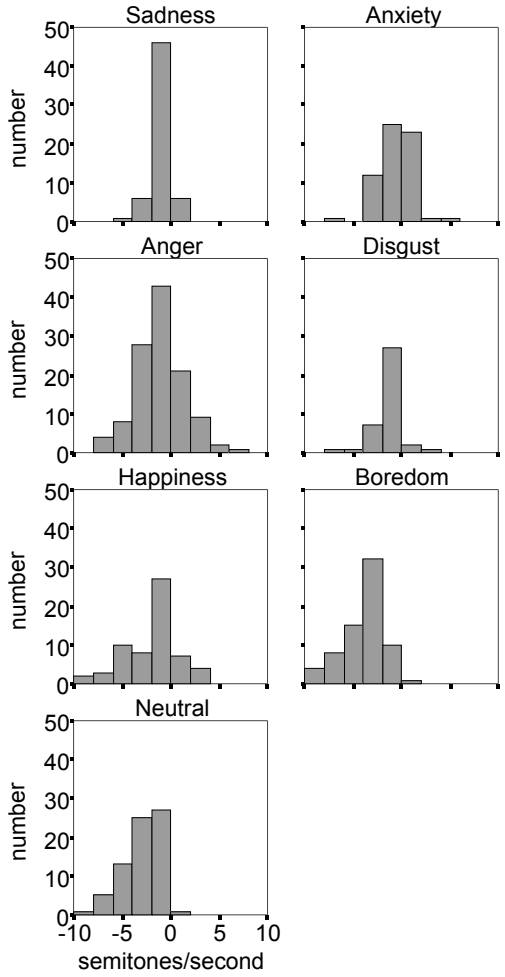

Figure 1: *Histograms of global trend for each emotion and neutral speech*.

Anxiety: Global trend for anxious utterances comes with a mean of –0.8 semitones per second. Even though the mean value for anxiety does not differ significantly from the values for anger, disgust and sadness, the variation differs highly. Whereas anger reveals the greatest variation, sadness and disgust show only minimal variation. For anxiety the variation is moderate. Most regression lines in anxious utterances are straightforward. The rest has a slight downward trend.

If one assumes that in neutral speech physiological factors cause the declination of the fundamental frequency over the duration of an utterance and that the mean global trend in neutral speech is –3 semitones per second, one could ask how it is possible to disable these constraints in emotional speech. The explanation which is acceptable for sadness (that based on the lower arousal, air consumption decreases, and therefore the global trend decreases) cannot hold for anxiety. In anxious speech the air consumption is expected to be higher than in neutral speech because of high arousal, and to result in a higher fundamental frequency. In addition, speakers often use breathy voice to express anxiety. This increases air consumption even more. In the recordings of the sentences expressing anxiety, speakers often make short pauses where inhalation is clearly audible. If there are no pauses, inhalation noises are audible after the last syllable. Thus the reason for a nearly straightforward global trend in anxious utterances can be found in the specific kind of breathing used, rather than in consciously generated pitch accents (such as was found for anger).

Disgust: The distribution of global trend values for disgust is very narrow. The typical global trend is slightly downwards. The gradient is smaller than in neutral speech which again can be explained with the specific kind of speaking style. There are a lot of laryngealized speech parts. Laryngalization comes with very low tension of the vocal chords and very low subglottal pressure so the air consumption is small. Another possible explanation is that through the accentuation typical for disgust the normal downward trend is superimposed with great increasing and decreasing fundamental frequency movements and therefore partly abolished.

Boredom: The greatest downward trend was found in the bored productions of sentences. The gradient averaged –4 semitones per second and reaches maximum values of –10 semitones per second. This is a strong contrast to the minimal downtrend in sad utterances, which leads to the conclusion that the low arousal which both emotions have in common cannot be a correlate of the arousal dimension in the dimensional emotion model.

However, the strongly falling global trend in bored utterances corresponds well with the auditory impression. Speakers take a deep breath before they start to speak and begin the sentence with a very high fundamental frequency level. Over the course of the utterance the fundamental frequency undergoes a strong fall with clearly downstepped pitch accents. The strong lowering of fundamental frequency (with the same degree of loudness) together with an audible exhalation (high air consumption) probably evokes the impression of boredom by the listener. The German saying “the air is out” describes exactly the feeling of suddenly occurring boredom.

The strong negative global trend of bored utterances can be explained either by the active control of fundamental frequency progression by the speaker or, if one assumes declination to be a process inherent to speech production and not controllable by the speaker, by increased breathing activity

as part of the physiological changes associated with this specific emotion.

The observation of a negative correlation between utterance duration and gradient of regression (r=0.79 for bored utterances) supports the assumption that speakers let out a certain amount of air which they have to ration about the whole utterance. The correlation coefficients for the other emotions are noticeably smaller.

5. Conclusions

Global trend is a suitable parameter to describe some emotions, especially emotions with only low arousal like boredom and sadness. It is less suitable for emotions characterized by high arousal (e.g. anger, happiness) where there is a great variation in global trend among different utterances. Speakers are free to use their whole voice range, including strong increasing global trends of fundamental frequency, to transfer these emotions.

Thus it can be concluded that declination is not a constant factor in speech production, but that various factors (like emotional arousal) have strong influence on underlying mechanisms. If a speaker expends more vocal effort than in normal speech (which results in higher subglottal pressure as well as in higher laryngeal tension) he can produce fundamental frequency movements which partly or even completely overlay the downward trend.

6. References

[1] Banse, R., Scherer, K. R., 1996. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, vol. 70 (3), p. 614-636.

[2] Cohen, A., Collier, R., ´t Hart, J., 1982. Declination: Construct or Intrinsic Feature of Speech Pitch? *Phonetica* 39, p. 254-273.

[3] Cohen, A., ´t Hart, J., 1967. On the anatomy of intonation. *Lingua* 19, p. 177-192.

[4] Collier, R., 1975. Physiological Correlates of Intonation Patterns. *JASA* 58, p. 249-255.

[5] Collier, R., 1983. Physiological Explanations of F0 Declination. In: A. Cohen, M. V. D. Broecke (eds.) *Abstracts of the 10th ICPhS*, p. 440.

[6] Cooper, W. E., Sorensen, J. M., 1981. Fundamental Frequency in Sentence Production. New York, Springer.

[7] Frick, R. W., 1985. Communicationg Emotion: The Role of Prosodic Features. Psychological Bulletin, vol. 97 (3), p. 412-429.

[8] Gussenhoven, C., Rietveld, A., 1988. Fundamental frequency declination in Dutch: testing three hypothesis. *Journal of Phonetics* 16, p. 355-369.

[9] Ladd, R.D., 1984. Declination: A Review and Some Hypotheses. In: *Phonology Yearbook* 1, pp. 53-74.

[10] Lieberman, P., 1967. Intonation, perception, and language. Cambridge, MIT Press.

[11] Maeda, S., 1976. A Characterization of American English Intonation. Ph.D. thesis, MIT Cambridge.

[12] Morton, K., Tatham, M., Lewis, E., 1999. A new intonation model for text-to-speech synthesis. In: *Proceedings of the 13th ICPhS*, San Francisco. p. 85-88.

[13] Murray, I. R., Arnott, L., 1993. Towards the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *JASA*, vol. 93 (2), p. 1097-1108.

[14] Ohala, J. J., 1990. Respiratory activity in speech. In W. J. Hardcastle & A. Marchal (eds.), *Speech production and speech modelling*. Dordrecht: Kluwer. p. 23-53.

[15] Paeschke, A.; Kienast, M.; Sendlmeier, W.F., 1999. F0-Contours in Emotional Speech. In: *Proceedings of the ICPhS99*, San Francisco, p. 929-932.

[16] Paeschke, A.; Sendlmeier, W. F., 2000. Prosodic Characteristics of Emotional Speech: Measurements of Fundamental Frequency Movements. In: *Proceedings of the ISCA Workshop on Speech and Emotion*, Textflow, Belfast, Northern Ireland, p. 75-80.

[17] Paeschke, A., 2003. Prosodische Analyse emotionaler Sprechweise, Ph.D. thesis TU Berlin and series „Mündliche Kommunikation“ 1, Logos, Berlin.

[18] Pierrehumbert, J., 1979. The perception of fundamental frequency declination. *JASA*, 66, p. 363-369.

[19] Pierrehumbert, J., 1980. The Phonology and Phonetics of English Intonation. Ph.D. thesis, MIT Cambridge.

[20] Pike, L., 1945. The Intonation of American English. University of Michigan Press, Ann Arbor, Michigan.

[21] Scherer, K. R., Banse, R., Wallbott, H. G., 2001. Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology*, vol. 32 (1), p. 76-92.

[22] Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., Gielen, S., 2001. Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis. In: Proceedings of Eurospeech, Aalborg, vol. 1, p. 87-90.

[23] Sendlmeier, W. F., 2001. Phonetische Variation als Funktion unterschiedlicher Sprechstile. In: Elektronische Sprachsignalverarbeitung, w.e.b., Dresden, p. 23-35.

[24] Sendlmeier, W. F., 2002. Stimmliche und phonetische Manifestation emotionaler Sprechweise. In: H. Geißner (ed.) Stimmkulturen, Röhrig Universitätsverlag, St. Ingbert, p. 39-49.

[25] Strik, H., Boves, L., 1995. Downtrend in F_0 and P_{sb}. *Journal of Phonetics*, 23, p. 203-220.

[26] Swerts, M., Strangert, E., Heldner, M., 1996. F0 Declination in Read-Aloud and Spontaneous Speech. In: *Proceedings of the ICSLP*, Philadelphia, vol. 3, p. 1501-1504.

[27] Tatham, M., Lewis, E., 1999. An advanced intonation model for synthesis. In: *Sixth European Conference on Speech Communication and Technology*, ESCA, p. 1871-1874.

[28] ´t Hart, J., Collier, R., 1979. On the interaction of accentuation and intonation in Dutch. In: *Proceedings of the 9th ICPhS*, Copenhagen, p. 395-402.

[29] Umeda, N., 1982. F0 Declination is situation-dependent. *Journal of Phonetics*, vol. 10, p. 279-290.

[30] Vaissière, J., 1983. Language-independent prosodic features. In: A. Cutler, D. R. Ladd (eds.) *Prosody: Models and Measurements*, Springer, Berlin, p. 53-66.