

The effects of emotional state on fundamental frequency

Louise Probst, Angelika Braun

University of Trier, Germany
s2lodirk@uni-trier.de

ABSTRACT

Although the role of fundamental frequency in the expression of emotion is a widely researched topic, individual results vary and are sometimes even contradictory. One potential reason for this are varying degrees of emotion being studied. The present contribution addresses this issue by taking into account three different levels of six emotions with neutral stimuli serving as a reference. Six native speakers of German were asked to portray the said emotions in three degrees: low, medium, and extreme. For *joy*, *hot anger* and *sadness*, results largely confirm the predictions based on previous research, but expectations are not met for *fear* and *cold anger*. Speakers follow a linear model for the expression of degrees of emotionality for most emotions, best represented by the expression of *disgust*. The present study demonstrates that it is highly advisable to consider varying degrees of emotionality when studying the effect of emotional state on vocal parameters.

Keywords: emotion, fundamental frequency

1. INTRODUCTION

Various acoustic parameters have been studied in relation to the encoding of emotions. There is broad consensus on fundamental frequency (f0) being an important feature in distinguishing emotions in speech. Even though f0 is probably the parameter which has been most frequently studied (and seems to be the most salient perceptually), individual results concerning the f0 values for emotions differ.

An exhaustive overview would by far exceed the scope of this contribution. This account will therefore focus on the findings for German and a few meta studies in addition. As far as results are concerned, the findings on f0 for both *joy* and *anger* are consistent across studies: both emotions cause speakers to produce a higher mean f0, greater f0 variability and greater f0 range than *neutral* utterances do. For *fear*, authors also report these parameters to increase as compared to *neutral* samples, albeit with different magnitudes [10; 13]. A decrease in mean f0 [2; 5; 7; 9-13] and in f0 variability [11] is found for *sadness*, but contradicting results exist for f0 range: Whereas

Klasmeyer [9] reports a greater f0 range as compared to *neutral*, Paeschke [11] finds a smaller one.

These so-called “basic emotions” (*joy*, *sadness*, *fear* and *anger*) are well-researched and are said to occur universally ([4]; for an overview also see [14] and [6]). Deviating somewhat from the concept of basic emotions, Scherer [13] (as well as Banse/Scherer [2]) make a case for the distinction between *hot anger* and *cold anger*, based on differences in the externalisation of these different facets of that emotion. Acoustic findings support this distinction and indicate that *hot* and *cold anger* are externalised by very different acoustic cues. Therefore, both facets of this emotion were included in the present study. A study by Braun/Heilmann [5] addresses dubbed speech in an intercultural setting. For their German speakers, results confirm those cited above. They consider *hot* and *cold anger* as different emotions, and report higher mean f0, variability and range for *hot anger*, and lower mean f0 (no change for variability and range) for *cold anger*.

An emotion which has only rarely been studied is *disgust*. It seems to be a “difficult” emotion as far as instructions to the speakers during the recording procedure are concerned [2]. Paeschke [11] finds a higher mean f0 and a wider f0 range for *disgust* as compared to the *neutral* samples, whereas f0 variability is very similar to them. In their overviews of studies on emotion in speech Banse/Scherer [2] and Pittam/Scherer [12] report inconsistent results, some authors describing increases and others decreases for mean f0 as compared to *neutral*.

Little attention has so far been devoted to various degrees of emotionality. Scherer attempts to “differentiate variants of a similar type of emotion” [13, p. 147], thus distinguishing e.g. between *fear* and *anxiety*, *joy* and *happiness*. But to the present authors' knowledge, there has never been an attempt to study varying degrees of the same emotion from an acoustic phonetic point of view. Therefore, a different approach was chosen for the experimental design in the present study, i.e. to instruct the speakers to portray three degrees of emotional expression (low, medium and extreme). The main research question is thus

whether the acoustic representations of the various degrees of emotionality are linear or whether they may be categorically different. This research might at the same time help to explain differing or even contradictory results of previous studies.

Based on the findings of the studies mentioned above and the predictions made by Banse/Scherer [2], an increase in mean f_0 as compared to *neutral* can be expected for the emotions *joy*, *fear*, *hot anger* and *cold anger*, and lower values than *neutral* for *sadness*. *Disgust* may render different results for the three degrees of emotion. Melodiousness (measured in terms of standard deviation of f_0) can be expected to increase for *hot anger*, *fear* and *joy*, and decrease for *cold anger* and *sadness*, making the latter sound more monotonous.

2. MATERIAL AND METHOD

The aim of this study was to work with the voices of speakers who are actors not solely capable of producing emotions “on the spot”, but also of using their voices only (i.e. without gestures, facial expressions and body language). The six professional speakers (native German speakers; 3 female, 3 male) chosen had many years of experience in acting, radio plays, dubbing and voice-overs for TV. They were asked to produce five nonsense sentences (existing words and grammatically correct, but no plausible meaning), all in six different basic emotions and in a *neutral* speaking style. For the emotions, the task was to express three degrees: low, medium and extreme (resulting in 90 utterances per emotion and a total of 570 utterances). The emotions covered were: *fear*, *disgust*, *joy*, *sadness*, *hot anger* and *cold anger*. No further instructions were given to the speakers, and none of them had any trouble or questions concerning the task. Recordings were made in a professional studio of the WDR radio station (Westdeutscher Rundfunk) using a Neumann U 87 and a DHD-RM4200 preamplifier at a sampling rate of 48kHz and 16 bit amplitude resolution. Speakers decided for themselves which samples were “best” and could repeat them until they were satisfied with the result.

Mean fundamental frequency (mean f_0), standard deviation (SD) and range (max f_0 – min f_0) were analysed for all five sentences. In extreme emotional speech a very wide range of pitch values can be expected. In order to avoid any errors in the automatic pitch extraction, the analysis was carried out manually, using PRAAT [4]. Statistical analyses for differences between *neutral* and emotional samples regarding each acoustic

parameter were tested by means of a paired Student's t-test, significances in the present study are based on $\alpha=0.05$ unless mentioned otherwise.

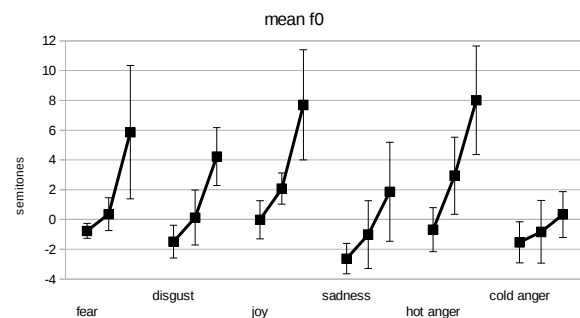
3. RESULTS

Emotionally loaded fundamental frequencies (f_0) were arranged relative to the mean values of the respective speaker's *neutral* speaking style. They were converted to (musical) semitones (ST) in order to facilitate between-speaker comparisons. A difference of one semitone corresponds to a f_0 change in Hertz of roughly six percent.

3.1. Mean f_0

Figure 1 provides an overview of mean f_0 behaviour for all speakers, for all emotions. The value for *neutral* stimuli is represented by zero. Mean f_0 values are lowest for low *sadness* and highest for extreme *hot anger*. Male speakers all show their highest mean f_0 for extreme *hot anger*, while female speakers portray extreme *joy* (and in one case extreme *fear*) with the highest f_0 .

Figure 1: Mean f_0 (and SD) for all speakers, all emotions and degrees. Values in semitones.



Increasing degrees of emotionality are in fact implemented by increasing mean f_0 for the portrayal of emotions. For *fear*, *disgust*, *joy* and *hot anger*, speakers increase their average f_0 across the three degrees of emotion (from low to extreme). They also follow this trend for *sadness*. Compared to *neutral*, all speakers mark low *sadness* with a significantly lower f_0 . For *fear*, *disgust*, *joy* and *hot anger*, the extreme degree of emotion exhibits a significant increase in mean f_0 .

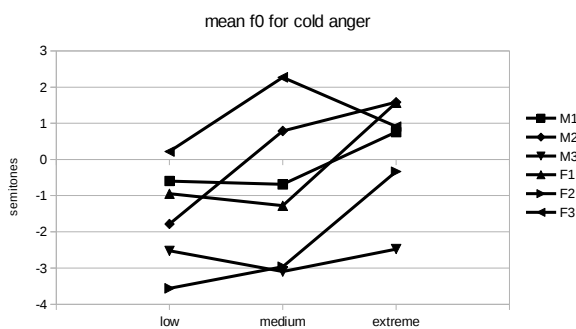
Low and medium *sadness* is expressed by a decrease in f_0 for most speakers, but four (two of them male, two female) show an increase for extreme *sadness*, and two male speakers show the highest difference (significant decrease) to *neutral* in the medium degree of this emotion. Table 1 gives an overview for trends for all speakers, for increasing and decreasing f_0 .

Table 1: F0 common ground for all speakers, * marks significant difference to neutral, blank areas denote different inter-speaker behaviour.

	low	medium	extreme
fear	decrease		increase*
disgust	decrease		increase*
joy		increase	increase*
sadness	decrease*	decrease	
hot anger		increase	increase*
cold anger	decrease		

Cold anger, however, seems a special emotion, and its portrayal to be largely speaker-specific. No clear trend is observed in terms of linearity for the three degrees of emotion.

Figure 2: Mean f0 for cold anger for all speakers.



3.2. F0 variability – Standard deviation (SD)

F0 variability (measured as standard deviation of f0) can be taken as an indication of the melodiousness/monotony of a given speaker's voice. Most interesting is here to look at differences of the emotional values in comparison to the *neutral* version of the respective speaker to get an impression of how they change the melodiousness of their voice in order to portray a specific emotion.

All speakers exhibit a decrease in SD for low and medium *fear*, but except for one female speaker, all mark extreme *fear* with a wider f0 variability. *Sadness* is portrayed with a smaller SD by all but one speaker, making it sound more monotonous. However, this decrease is not linear concerning the three degrees of emotionality, speakers seem to handle this emotion categorically. *Joy* and (for most speakers) *disgust* are marked with increasingly wider f0 variability across the three steps, and most speakers handle *hot anger*

and *cold anger* similarly. *Hot anger* and – for most speakers – *cold anger* are marked by an increase in SD for medium and extreme degree of emotionality (extreme *hot anger* significant for all speakers for $\alpha=0.01$). Table 2 gives an overview of the standard deviation of f0 and significances for all speakers, as compared to their respective *neutral* recording.

Table 2: SD common ground for all speakers, * marks significant difference to neutral, blank areas denote different inter-speaker behaviour.

	low	medium	extreme
fear	decrease		increase
disgust	decrease	increase	increase
joy	increase	increase*	increase*
sadness	decrease	decrease	
hot anger	increase	increase	increase*
cold anger		increase	increase

For all speakers, f0 variability is largest for extreme *joy* and extreme *hot anger* (both significant for $\alpha=0.01$ for all speakers).

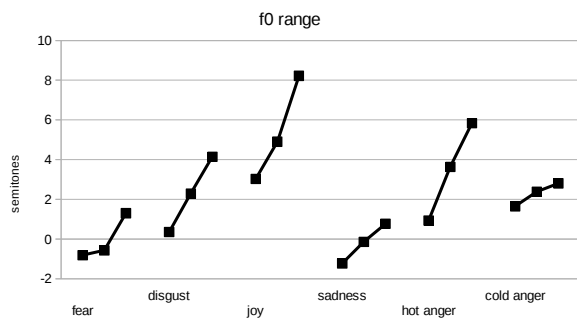
A pattern (linear increase) in SD can be observed for *disgust*, *joy* and mostly for *fear*, *hot anger* and *cold anger*. *Sadness* is handled rather categorically concerning the SD.

3.3. F0 range

F0 ranges – as the distance between highest and lowest measured f0 – were calculated for each utterance separately. Mean ranges per degree per emotion were determined and also compared to the mean f0 range of the *neutral* samples. F0 ranges for the emotions extend from -4.9 semitone for medium and extreme *fear* up to 12 semitones wider than *neutral* for extreme *joy* (which corresponds to a musical octave).

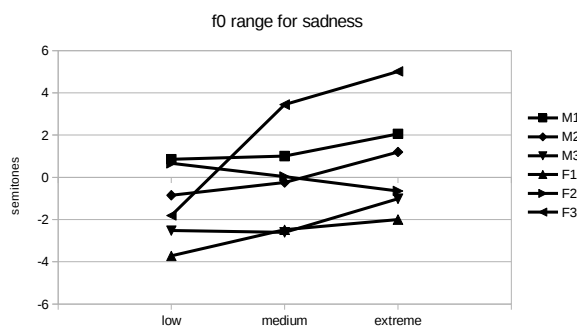
An extreme increase in f0 range is portraying *joy*, reaching significance for the medium and the extreme degree (the latter for $\alpha=0.01$). *Disgust*, *hot anger* and for most speakers *cold anger* are also portrayed with increased f0 range. Also, speakers increase f0 range further over the three degrees of emotionality, again following a linear pattern. Figure 3 provides an overview of f0 range and shows common ground for all speakers.

Figure 3: F0 range for all speakers, all emotions and degrees.



Low *sadness* is clearly marked by a smaller f0 range than *neutral* by all speakers, but they don't use varying f0 range much to distinguish between the degrees of emotionality. No clear trend is observed in terms of linearity for the three degrees of emotionality, as is shown in Figure 4.

Figure 4: F0 range for sadness.



4. DISCUSSION

As the data base of the present study is relatively limited, all results have to be handled with caution. Concerning the three degrees of emotion, it becomes clear from the data that no equidistance is given between low to medium and medium to extreme emotion. Predictions and findings of previous studies [2; 5; 9-13] are confirmed by the results for *joy* and *hot anger*, and mostly for *sadness* (rise or lowering for f0 range here seems to be speaker-specific). Speakers handle *fear* more individually than predicted, they differ in their changes of f0 variability and f0 range (increase or decrease). Predictions do not hold for *cold anger*. All speakers show by far lower values for *cold* than for *hot anger*, and mark it by greater standard deviation and greater range. The distinction between *hot anger* and *cold anger* as suggested by Scherer [13] is confirmed, as all speakers clearly

mark them as different emotions. *Disgust* seems (at least in the given setting) not as complex an emotion as predictions suggest. All speakers handle it with increasing mean f0 over emotion steps.

Speakers do follow a linear pattern for the three degrees of emotionality, best represented by *disgust* and *joy*. This linear model also holds true for *fear* and *hot anger*, though not for all speakers. The degrees for *cold anger* and *sadness* seem to be handled rather categorically instead.

5. CONCLUSION

In order to portray emotions, speakers use modulations of their fundamental frequency. Extreme emotions come with a more extreme change of f0 as compared to *neutral* speaking style. Though some externalisations are similar for all speakers, there are some emotions for which portrayals differ greatly (as *fear*, *disgust* and *cold anger*). In most cases, extreme emotions are marked more consistently than low or medium emotions. Speakers tend to mark the three degrees of emotionality by a linear pattern, increasing the measured parameters step by step.

Different or even contradicting results in previous research might be due to difference in degree of emotionality. Most authors give their speakers clear instruction on how to portray a specific emotion (as detailed descriptions, scenes [15]), elicit them by images [1] and videos or evoke them by specially designed tasks [3; 8]. However, the degree of emotion is seldom included (and if so, it is subsumed under different emotional labels), leaving the decision to the intention of the speaker. The findings of the present study demonstrate that it is highly advisable to consider varying at least a lesser and a higher degree of emotionality when studying the effect of emotional state on vocal parameters.

6. REFERENCES

- [1] Abelin, Å. 2005. Change of Perception of Emotions in Multimodal and Crosslinguistic Settings. *Proc. ISCA Workshop on Plasticity in Speech Perception*, London, 110-113.
- [2] Banse, R., Scherer, K. R. 1996. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* 70 (3), 614-636.
- [3] Batliner, A., Fischer, K., Huber, R., Spilker, J., Noth, E. 2000. Desperately seeking emotions or: actors, wizards, and human beings. In: *Proc. ISCA Workshop on Speech and Emotion*, Newcastle, Northern Ireland, 195-200.
- [4] Boersma, P., Weenink, D. Praat: doing phonetics by computer <http://www.praat.org>
- [5] Braun, A., Heilmann, C. 2012. *SynchronEmotion*. Frankfurt a. M.: Peter Lang.
- [6] Cowie, R. 2000. Describing the emotional states expressed in speech. *Proc. ISCA Workshop on Speech and Emotion*, Newcastle, Northern Ireland, 3-8.
- [7] Ekman, P. 1992. An Argument for Basic Emotions. *Cognition and Emotion* 6, 169-200.
- [8] Kehrein, R. 2002. *Prosodie und Emotionen*. Tübingen: Niemeyer.
- [9] Klasmeyer, G. 1999. *Akustische Korrelate des stimmlich emotionalen Ausdrucks in der Lautsprache*. Frankfurt a. M.: Hector.
- [10] Murray, I. R., Arnott, J. L. 1993. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *J. Acoust. Soc. Am.* 93(2), 1097-1108.
- [11] Paeschke, A. 2003. *Prosodische Analysen emotionaler Sprechweise*. Berlin: Logos.
- [12] Pittam, J., Scherer, K. R. 1993. Vocal Expression and Communication of Emotion. In: Lewis, M., Haviland, J. M. (eds.), *Handbook of Emotions*. New York/London: Guilford Press, 185-197.
- [13] Scherer, K. R. 1986. Vocal affect expression: A review and a model for future research. *Psych. Bulletin* 99 (2), 143-165.
- [14] Stein, N., Oatley, K. 1992. Basic Emotions: Theory and Measurement. *Cognition and Emotion* 6, 161-168.
- [15] Williams, C. E., Stevens, K. N. 1972. Emotions and Speech: Some Acoustical Correlates. *J. Acoust. Soc. Am.* 52(4), 1238-1250.