

COVER PAGE

Document downloaded by @DAEL

Fri Nov 27 14:38:17 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

A Note on Acoustic Features in Pitch Contours for Discrimination of Happiness and Anger, Mirjana Rajković, Slobodan T. Jovičić, Dorde T. Grozdić, Svetlana Zdravković and Miško Subotić, Acta Acustica **vol. 104** (Number 2), 2018, pp. 369-372

DOI
https://doi.org/10.3813/AAA.919179

Acta Acustica united with Acustica
Vol. 104 (2018) 369 – 372

DOI 10.3813/AAA.919179

A Note on Acoustic Features in Pitch Contours for Discrimination of Happiness and Anger

Mirjana Rajković[1)], Slobodan T. Jovičić[1,2)], Đorđe T. Grozdić[1,2)], Svetlana Zdravković[3)], Miško Subotić[2)]

[1)] Department of Telecommunication, University of Belgrade, School of Electrical Engineering, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia. mink@eunet.rs, jovicic@etf.rs, djordje grozdic@gmail.com

[2)] Life activities advancement center, Institute for experimental phonetics and speech pathology, Gospodar Jovanova 35, 11000 Belgrade, Serbia. ifp2@ikomline.net

[3)] Institute of Mental Health, Palmotićeva 37, 11000 Belgrade, Serbia. svetlana.zdravkovic@imh.org.rs

Summary

The intonation pattern of speech consists of many short intonation contours, so-called individual intonation contours (IICs). This paper is concerned with the acoustic features of IICs for five emotional categories: anger, fear, sadness, happiness and neutral. The acoustic features are: mean duration of IICs, mean frequency dynamic of IICs (pitch-Fo variation), mean duration between the Fo maximum and the minimum inside IICs, relative number of IICs with rising shape and *EAC* (emotional articulatory changes), defined on a raising part of IICs. The results have shown that the *EAC* is the most effective feature in the discrimination between happiness and anger.

PACS no. 43.71.Bp, 43.70.Fq

1. Introduction

Beside basic linguistic information, the speech also transmits a variety of nonlinguistic data such as: quality of voice, quality of articulation, gender, age, emotions, intentions, etc [1, 2, 3]. Recently, the analysis of emotions expressed in speech has become a very promising topic of the investigation. Such results are important for application in the human-computer communication, particularly in the automatic recognition of emotion in speech and the synthesis of emotional speech [4, 5, 6]

Previous studies have shown that naive listeners are usually able to determine the emotional content of the utterances simply by hearing the speech signal. The 'Feeltrace', as an instrument for recording perceived emotions, demonstrated a very successful discrimination of emotions in the arousal-valence two-dimensional space [7, 8]. For instance, anger and happiness are at the opposite sides of the valence scale (negative–positive), but both are on the opposite side in regard to sadness and boredom across the arousal scale (active–passive). In the automatic detection of emotions in speech based on acoustic features, some emotions can be clearly discriminated, but some can be confused with others. The results presented in literature showed that sadness was often confused with fear or neutral emotion, happiness was frequently confused with anger, while sadness was almost never confused with anger or happiness [9, 10]. However, there is the evidence in the literature that the greatest confusion exists between the emotions of happiness and anger [11, 12, 13].

Received 3 August 2016,
accepted 2 February 2018.

Investigations of the features in speech signal, that are the symptom of the emotion, indicated the intonation patterns as the most important ones [14, 15, 16, 17]. A number of the intonation pattern features related to the pitch contour Fo(t) were analyzed in the literature, such as: Fo average, Fo maximum and minimum, Fo standard deviation, Fo onset, Fo offset, dynamic of pitch contour (Fo range), pitch contour shape (rising and falling slope of the overall contour shape), features based on Fo contours in syllables, Fo contour declination, linear regression coefficients [18, 6]. Those features are more or less correlated with some of the emotions, but there still remains a confusion in discrimination between anger and happiness. The results presented in literature [5, 6] revealed that the perceptual discrimination of these two emotions was more powerful than automatic discrimination based on any set of acoustic features.

This paper describes two preliminary experiments organized to explore possible contribution of some acoustic features on the individual intonation contour (IICs) for improving the automatic recognition of emotions in speech. The focus was put on the emotions of happiness and anger. In the first experiment we analyzed four features: ΔFo, Tc, ΔTc and S^+ in five emotional categories: anger, fear, sadness, happiness and neutral. The results of this experiment served as the basis for the second experiment where we defined a new acoustic feature *EAC* (emotional articulatory changes) and applied it for the discrimination between emotions: anger and happiness.

2. Experiment I

2.1. Nature of IICs

This experiment is organized to explore some basic features of individual intonation contours (IICs). We introduced IICs as a new concept which represented a continuous segments of intonation contour between two interruptions. The interruptions arose during unvoiced segments in speech or during pauses in voice between two speech segments.

Intonation contour for the same sentence is very diverse among speakers even for the same speaker depending on their style of speaking. The illustration of intonation contours and their potential differences is presented in Figure 1a. The figure shows the intonation contours in the Serbian sentence "Krećemo za pet minuta" (We start in five minutes) pronounced in anger, happiness and neutral manner by the same speaker. The duration of contour is the shortest for the neutral emotion and the longest for happiness. The average pitch for the neutral emotion is the lowest, while for anger and happiness it is higher and with comparable levels. The number of IICs is different in intonation contours: for neutral emotion there are four IICs, for anger there are five IICs and for happiness there are six IICs. Within marked area it could be seen that IIC for anger has the falling slope, in comparison to the rising slope of IIC for happiness and the falling/rising slopes of IIC for the neutral emotion.

This last observation shows that the IIC, as a constituent of intonation at the level of a sentence, contains features that could be informative for recognizing a particular emotion or for discriminating between emotions (for instance, feature S^+ that explained in the next section).

2.2. Feature of IICs

In this experiment a hypothesis was introduced that some of IICs features were correlated with some emotions. For quantitative characterization of IICs we defined three new features:

© S. Hirzel Verlag · EAA

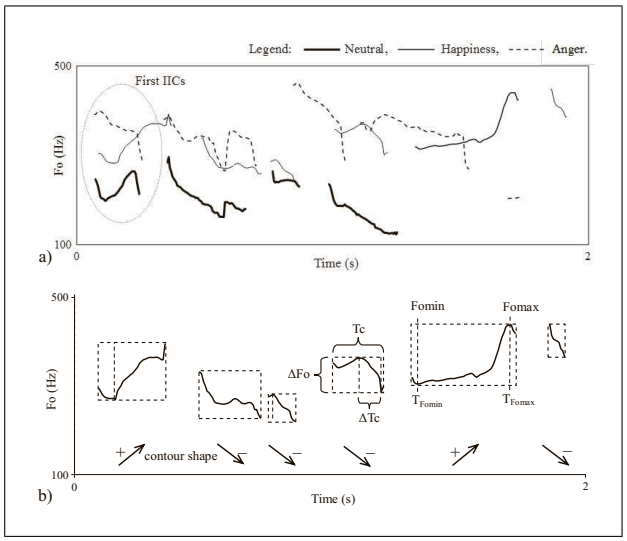


Figure 1. (a) Intonation contours in Serbian sentence "Krećemo za pet minuta" (We start in five minutes) pronounced in anger, happiness and with neutral emotion (marked area – First IICs – emphasized the differences in shapes of IICs); (b) definition of the features for IICs (example for happiness from Figure 1a).

ΔFo (Hz) – the difference between the maximum and the minimum of Fo in IIC, i.e. the dynamic of IIC in frequency domain (ΔFo = Fomax −Fomin),

Tc (ms) – the duration of IIC,

ΔTc (ms) – the distance between positions in time of Fomax and Fomin in IIC ($\Delta Tc = |T_{Fomax} - T_{Fomin}|$).

Beside these three features of IICs the following feature, defined for the entire sentence, was also introduced: S^+ – the relative number of IICs with rising shape (IIC^+) within the sentence, defined as $S^+ = n/N$, where n is the number of IIC^+ and N is the total number of IICs in the sentence. The rising shape IIC^+ is defined as IIC with first Fomin and than Fomax in time.

The illustration of these features is presented in Figure 1b (intonation contour for happiness from Figure 1a).

2.3. Experiment organization

The passage of a 79 words long speech was used in this analysis. The passage consists of nine sentences. It was taken from the Serbian emotional speech database GEES [19, 20]. In this database, six actors – three men and three women from the Faculty of Dramatic Arts, University of Belgrade, pronounced the same text with five emotions: anger, happiness, sadness, fear and neutral.

The software Praat [21] was used for speech signal analysis and for measuring four specified features. The features were automatically extracted from the F0 contour using some built-in functions.

2.4. Results

Figure 2 presents the average values and standard deviations of features Tc, ΔTc, ΔFo and S^+ for all five emotions and for both genders. The features Tc and ΔTc show consistent behavior and they were not significantly changed with emotions, except for the emotion of fear. The feature ΔFo is not consistent for males and females over emotions, while feature S^+ discriminates the emotion of happiness from other emotions for all speakers.

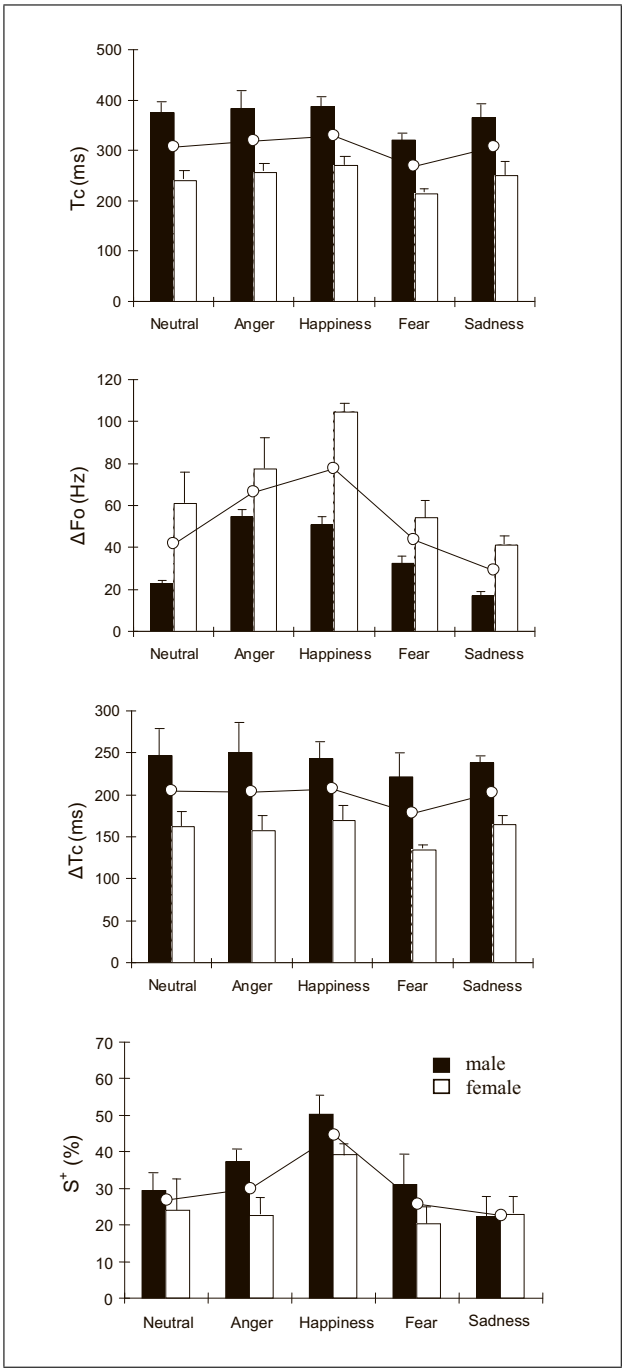


Figure 2. Distribution of features Tc, ΔTc, ΔFo and S^+ for all emotions and speakers (empty circles represent average values for all speakers, error bars represent standard deviation).

The results of statistical analysis for each feature with two factor ANOVA (factors are gender and emotion) are presented in Table I. There is a statistically significant difference between genders for all four features, as well as for emotions except for feature ΔTc (p=0.201). The interaction between the two factors indicates a statistically significant difference only for feature ΔFo (p=0.031).

Bonferroni post hoc test showed that for feature Tc existed a statistically significant difference only between the emotions of fear and anger (p = 0.018) and between fear and happiness (p = 0.003). For the feature ΔTc there was not a statistically significant difference between the emotions. For the feature ΔFo

Table I. The results for two factor ANOVA analysis.

Features	Factors	ANOVA
Tc	Gender	F(1,20) = 174.45, p<0.001
	Emotion	F(4,20) = 5.339, p=0.004
	Gender * Emotion	F(4,20) = 0.264, p=0.898
ΔTc	Gender	F(1,20) = 101.991, p<0.001
	Emotion	F(4,20) = 1.651, p=0.201
	Gender * Emotion	F(4,20) = 0.213, p=0.928
ΔFo	Gender	F(1,20) = 91.959, p<0.001
	Emotion	F(4,20) = 27.996, p<0.001
	Gender * Emotion	F(4,20) = 3.316, p=0.031
S^+	Gender	F(1,20) = 9.303, p=0.006
	Emotion	F(4,20) = 8.297, p<0.001
	Gender * Emotion	F(4,20) = 0.979, p=0.441

Table II. The results for two factor ANOVA analysis.

Feature	Factors	ANOVA
EAC	Gender	F(1,12) = 0.129, p=0.726
	Emotion	F(4,20) = 6.773, p=0.011
	Gender * Emotion	F(4,20) = 0.266, p=0.771

existed a statistically significant difference between the emotions of anger and happiness and other three emotions ($p < 0.05$), but not between anger and happiness ($p = 0.404$), and that for feature S^+ existed a statistically significant difference between emotion happiness and all other emotions ($p < 0.05$). Since our primary intention with this investigation was a discrimination between the emotions of happiness and anger, the results indicated that the rising slope of IICs pointed to information potentially important for discrimination between these two emotions. With this knowledge we performed a second experiment where we defined a new feature and applied it to emotions of happiness and anger in comparison to neutral emotion, in order to evaluate its effectiveness in discriminating these two emotions.

3. Experiment II

3.1. New feature definition

More detailed insight on IICs in Figure 1 shows that their structure is not simple (only rising or falling slope), but that it is more complex. As the illustration, we extracted from Figure 1a the first two IICs for happiness and anger and that is presented in Figure 3. It could be seen that each IIC has one or more rising and/or falling subsegments. In our discussion we focused on subsegments with large raising slope for happiness in the first IIC and large raising slope for anger in the second IIC, as the most characteristic subsegments for these two emotions. In speech pronounced in anger, this subsegment is steeper than this one pronounced in happiness. As a consequence, the time interval Δt_r, that represents duration of raising subsegment, is significantly shorter in anger than in happiness. On the other hand, pitch variation in raising subsegments $\Delta \mathrm{Fo}_r$ in both emotions is similar. In terms of articulatory changes, the anger makes an abrupt articulatory changes.

In our preliminary tests we found the product of $\Delta \mathrm{Fo}_r$ and Δt_r very useful in discriminating the emotions of happiness and anger. To support these findings, we defined a new feature *EAC*

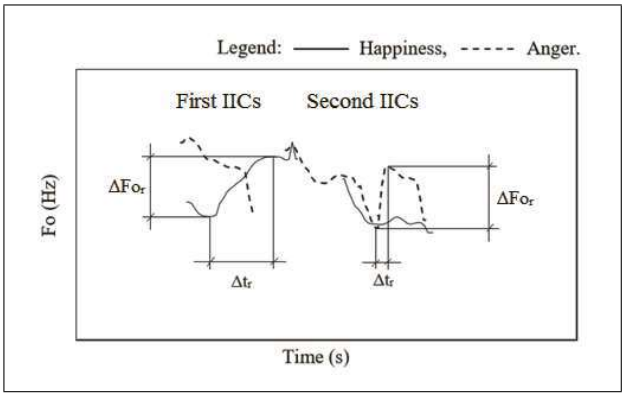


Figure 3. Illustration for definition of the feature *EAC*.

(emotional articulatory changes) and proposed it's model in the form of accumulated and normalized product $\Delta \mathrm{Fo}_r$ and Δt_r as

$$EAC = \left(\sum_{i=1}^{M} \Delta \mathrm{Fo}_{ri} \cdot \Delta t_{ri} \right) \frac{1}{\overline{\mathrm{Fo}} T}, \tag{1}$$

where $\overline{\mathrm{Fo}}$ is the Fo averaged along verbal expression, T is the duration of a verbal expression and M is the number of rising subsegments in verbal expression. Normalization to $\overline{\mathrm{Fo}}T$ in (1) introduces equalization between subjects. This is necessary because different subjects, male or female, have different $\overline{\mathrm{Fo}}$.

3.2. Experiment organization

The following experiment was performed to test the hypothesis that *EAC* is a feature which can be used in the discrimination of the emotions of happiness and anger. As in the experiment I, a part of the GEES corpus was used: 30 long sentences uttered by six speakers in three basic emotions: anger, happiness and neutral speech [19]. The neutral speech was used as a reference in comparing emotions of anger and happiness. That made 180 sentences for each emotion, i.e. 540 speech statements in total. In this experiment we used a larger speech material compared to experiment I (30 sentences versus 9 sentences in the passage) in order to obtain a more robust statistical results. The parameters ΔFo and Δt_r were measured for each rising subsegment inside each IIC, according to the definition in Figure 3, and *EAC* was estimated for each sentence according of (1). The final value of *EAC* for each emotion was obtained by averaging over all sentences and speakers. The parameters ΔFo and Δt_r were automatically extracted from the F0 contour using some built-in functions in software Praat.

3.3. Results

The average results for all speakers are shown in Figure 4. The results demonstrate clear discrimination between happiness and other two emotions. ANOVA test (Table II) shows a statistically significant difference between the emotions [$F(2,10) = 6.773$, $p = 0.011$], but not between genders. This result indicates the fact that the feature *EAC* is robust to genders. In the next step Bonferroni post hoc test was performed and it showed a statistically significant difference between the emotions of happiness and anger ($p = 0.038$), and happiness and neutral speech ($p = 0.016$).

It is important to emphasize the difference in *EAC* values between happiness and anger (Figure 4). Example in Figure 3 supports the fact that large raising subsegments in both emotions are dominant in equation (1).

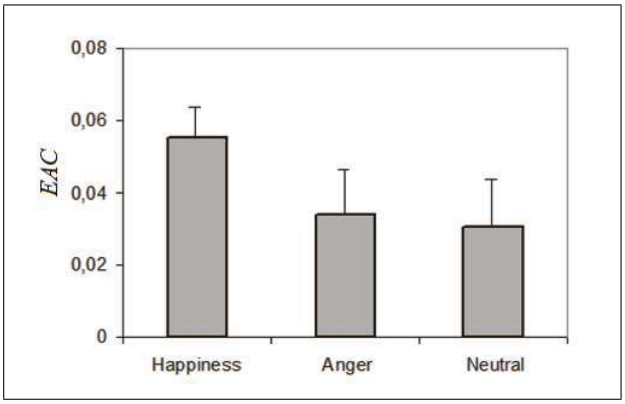


Figure 4. Results of experiment II.

4. Conclusion

This study examined the relative contribution of some individual intonation contour features (IICs) to the vocal expression of the following emotions: anger, happiness, sadness, fear and neutral speech, and especially to the discrimination between happiness and anger. The experimental results support the hypothesis that the introduced acoustic feature *EAC* (emotional articulatory changes), defined on raising subsegments of IICs, successfully discriminates between the positive emotion – happiness, and the negative emotion – anger, as two opposite emotions on the valence scale. On the other hand, the feature *EAC* supports the fact that man evokes more energy in terms of articulatory changes to make an abrupt change in anger comparing to a slower change in happiness. Also, results demonstrate that the feature *EAC* is very robust regarding to individual variations in expression of emotions. It is expected that feature *EAC* could be useful for real time implementation in human-computer communication systems. Finally, experiment II indicates that each emotion has a predominant feature, which distinguishes it from other emotions, and it would be reasonable to assume that this feature could have different weights for different emotions. This view is a matter of future investigation.

Acknowledgement

This research was supported by grants OI178027 and TR32032 from the Ministry of Education, Science and Technological Development of the Republic of Serbia.

References

[1] C. E. Williams, K. N. Stevens: Emotions and speech: some acoustic correlates. J. Acoust. Soc. Am. **52** (1972) 1238–1250.

[2] J. J. L. Mozziconacci: Speech variability and emotion: Production and perception. Ph.D. Thesis, Technical University, Eindhoven, 1998.

[3] H. Quast: Automatic recognition of nonverbal speech: An approach to model the perception of paraand extralinguistic vocal communication with neural networks. Thesis. University of Göttingen, 2001.

[4] I. R. Murray, J. L. Arnott: Towards a simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. J. Acoust. Soc. Am. **93** (1992) 1097–1108.

[5] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J. Taylor: Emotion recognition in human-computer interaction. IEEE Signal Process. Mag. **18** (2001) 32–80.

[6] M. E. Ayadi, M. S. Kamel, F. Karry: Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition **44** (2011) 572–587.

[7] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, M. Schröder: "feeltrace": An instrument for recording perceived emotion in real time. Proc. of the ISCA Workshop on Speech and Emotion, 2000, 19–24.

[8] S. T. Jovičić, M. Rajković, M. Đordević, Z. Kašić: Perceptual and statistical analysis of emotional speech in man-computer communication. 11th Int. Conf. SPECOM'2006, St. Petersburg, Russia, 2006, 409–414.

[9] R. Banse, K. R. Scherer: Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology **70** (1996) 614–636.

[10] S. Yildirim, M. Bulut, C. Lee, A. Kazemzadeh, C. Busso, Z. Deng, S. Lee, S. Narayanan: An acoustic study of emotions expressed in speech. Proc. of ICSLP, 2004, 2193–2196.

[11] M. Schröder: Speech and emotion research: An overview of research frameworks and a dimensional approach to emotional speech synthesis. Dissertation, Universität des Saarlandes, 2003.

[12] M. C. Sezgin, B. Gunse, G. K. Kurt: Perceptual audio features for emotion detection. EURASIP Journal on Audio, Speech, and Music Processing **16** (2012).

[13] A. Origlia, F. Cutugno, V. Galata: Continuous emotion recognition with phonetic syllables. Speech Communication **57** (2014) 155–169.

[14] S. J. Mozziconacci, D. J. Hermes: Role of intonation patterns in conveying emotion in speech. ICPhS'99, 1999, 2001–2004.

[15] T. Bänziger, K. R. Scherer: The role of intonation in emotional expressions. Speech Communication **46** (2005) 252–267.

[16] S. T. Jovičić, M. Rajković, M. Đordević, Z. Kašić: Prosodic characteristics of emotional speech: Analysis of dynamic behaviour of a group of acoustic features. Int. Conf. Forum Acusticum, Budapest, 2005, 211–214.

[17] E. Rodero: Intonation and emotion: Influence of pitch levels and contour type on creating emotions. Journal of Voice **25** (2011) e25–e34.

[18] P. N. Juslin, P. Laukka: Communication of emotions in vocal expression and music performance: Different channels, same code? Psychological Bulletin **129** (2003) 770–814.

[19] S. T. Jovičić, Z. Kašić, M. Đordević, M. Rajković: Serbian emotional speech database: design, processing and evaluation. Int. Conf. SPECOM'2004, St. Petersburg, Russia, 2004, 77–81.

[20] ELRA-S0385: Serbian emotional speech database (GEES). ISLRN: 462-780-920-598-3, http://catalog.elra.info/product_info.php?products_id=1277, Sept. 2016.

[21] Praat. Version 5.2.24, http://www.fon.hum.uva.nl/praat/, 2011.