23

Acoustic Patterns of Emotions

Branka Zei Pollermann and Marc Archinard
Liaison Psychiatry, Geneva University Hospitals
CH-1205 Geneva, Switzerland
vox-institute@swissonline.ch

Introduction

Naturalness of synthesised speech is often judged by to how well it reflects the speaker's emotions and/or how well it features the culturally shared vocal prototypes of emotions (Scherer, 1992). Emotionally coloured vocal output is thus characterised by a blend of features constituting patterns of a number of acoustic parameters related to F0, energy, rate of delivery and the long-term average spectrum.

Using the covariance model of acoustic patterning of emotional expression, the chapter presents the authors' data on: (1) the inter-relationships between acoustic parameters in male and female subjects; and (2) the acoustic differentiation of emotions. The data also indicate that variations in F0, energy, and timing parameters mainly reflect different degrees of emotionally induced physiological arousal, while the configurations of long term average spectra (more related to voice quality) reflect both arousal and the hedonic valence of emotional states.

Psychophysiological Determinants of Emotional Speech Patterns

Emotions have been described as psycho-physiological processes that include cognitions, visceral and immunological reactions, verbal and nonverbal expressive displays as well as activation of behavioural reactions (such as approach, avoidance, repulsion). The latter reactions can vary from covert dispositions to overt behaviour. Both expressive displays and behavioural dispositions/reactions are supported by the autonomic nervous system which influences the vocalisation process on three levels: respiration, phonation and articulation. According to the *covariance* model (Scherer *et al.*, 1984; Scherer and Zei, 1988; Scherer, 1989), speech patterns covary with emotionally induced physiological changes in respiration, phonation and articulation. The latter variations affect vocalisation on three levels:

1. suprasegmental (overall pitch and energy levels and their variations as well as timing);
2. segmental (tense/lax articulation and articulation rate);
3. intrasegmental (voice quality).

Emotions are usually characterised along two basic dimensions:

1. activation level (aroused vs. calm), which mainly refers to the physiological arousal involved in the preparation of the organism for an appropriate reaction;
2. hedonic valence (pleasant/positive vs. unpleasant/negative) which mainly refers to the overall subjective hedonic feeling.

The precise relationship between the physiological activation and vocal expression was first modelled by Williams and Stevens (1972) and has received considerable empirical support (Banse and Scherer, 1996; Scherer, 1981; Simonov *et al.*, 1980; Williams and Stevens, 1981). The activation aspect of emotions is thus known to be mainly reflected in the pitch and energy parameters such as mean F0, F0 range, general F0 variability (usually expressed either as SD or the coefficient of variation), mean acoustic energy level, its range and its variability as well as the rate of delivery. Compared with an emotionally unmarked (neutral) speaking style, an angry voice would be typically characterised by increased values of many or all of the above parameters, while sadness would be marked by a decrease in the same parameters. By contrast, the hedonic valence dimension, appears to be mainly reflected in intonation patterns, and in voice quality.

While voice patterns related to emotions have a status of symptoms (i.e. signals emitted involuntarily), those influenced by socio-cultural and linguistic conventions have a status of a consciously controlled speaking style. Vocal output is therefore seen as a result of two forces: the speaker's physiological state and socio-cultural linguistic constraints (Scherer and Kappas, 1988).

As the physiological state exerts a direct causal influence on vocal behaviour, the model based on scalar *covariance* of continuous acoustic variables appears to have high cross-language validity. By contrast the *configuration* model remains restricted to specific socio-linguistic contexts, as it is based on configurations of category variables (like pitch 'fall' or pitch 'rise') combined with linguistic choices. From the listener's point of view, naturalness of speech will thus depend upon a blend of acoustic indicators related, on the one hand, to emotional arousal, and on the other hand, to culturally shared vocal stereotypes and/or prototypes characteristic of a social group and its status.

Intra and Inter-Emotion Patterning of Acoustic Parameters

Subjects and Procedure

Seventy-two French speaking subjects' voices were used. Emotional states were induced through verbal recall of the subjects' own emotional experiences of joy, sadness and anger (Mendolia and Kleck, 1993). At the end of each recall, the subjects said a standard sentence on the emotion congruent tone of voice. The sentence was: 'Alors, tu acceptes cette affaire' ('So you accept the deal.'). Voices were digitally recorded, with mouth-to-microphone distance being kept constant.

The success of emotion induction and the degree of emotional arousal experienced during the recall and the saying of the sentence were assessed through self-report. The voices of 66 subjects who reported having felt emotional arousal while saying the sentence were taken into account (30 male and 36 female). Computerised analyses of the subjects' voices were performed by means of Signalyze, a Macintosh platform software (Keller, 1994). The latter provided measurements of a number of vocal parameters related to emotional arousal (Banse and Scherer, 1996; Scherer, 1989). The following vocal parameters were used for statistical analyses: mean F0, F0sd, F0 max/min ratio, voiced energy range. The latter was measured between two mid-point vowel nuclei corresponding to the lowest and the highest peak in the energy envelopes and expressed in pseudo dB units (Zei and Archinard, 1998). The rate of delivery was expressed as the number of syllables uttered per second. Long-term average spectra were also computed.

Results for Intra-Emotion Patterning

Significant differences between male and female subjects were revealed by the ANOVA test. The differences concerned only pitch-related parameters. There was no significant gender-dependent difference either for voiced energy range or for the rate of delivery: both male and female subjects had similar distributions of values regarding the rate of delivery and voiced energy range. Table 23.1 presents the F0 parameters affected by speakers' gender and ANOVA results.

Table 23.1 F0 parameters affected by speakers' gender

Emotions	F0 mean in Hz	ANOVA	F0 max/min ratio	ANOVA	F0 SD	ANOVA
anger	M 128; F 228		M 2.0; F 1.8		M 21.2; F 33.8	
		$F(1, 64) = 84.6$***		$F(1, 64) = 5.6$*		$F(1, 64) = 11.0$**
joy	M 126; F 236		M 1.9; F 1.9		M 22.6; F 36.9	
		$F(1, 64) = 116.8$***		$F(1, 64) = .13$		$F(1, 64) = 14.5$***
sadness	M 104; F 201		M 1.6; F 1.5		M 10.2; F 19.0	
		$F(1, 64) = 267.4$***		$F(1, 64) = .96$		$F(1, 64) = 39.6$***

Note: N = 66. *$p < .05$, **$p < .01$, ***$p < .001$; M = male; F = female.

As gender is both a sociological variable (related to social category and cultural status) and a physiological variable (related to the anatomy of the vocal tract), we assessed the relation between mean F0 and other vocal parameters. This was done by computing partial correlations between mean F0 and other vocal parameters, with sex of speaker being partialed out. The results show that the subjects with higher F0 also have higher F0 range (expressed as max/min ratio) across all emotions. In anger, the subjects with higher F0 also exhibit higher pitch variability (expressed as F0sd) and faster delivery rate. In sadness the F0 level is negatively correlated with voiced energy range. Table 23.2 presents the results.

Results for Inter-Emotion Patterning

The inter-emotion comparison of vocal data was performed separately for male and female subjects. A paired-samples t-test was applied. The pairs consisted of the same acoustic parameter measured for two emotions. The results presented in Tables 23.2 and 23.4 show significant differences mainly for emotions that differ on the level of physiological activation: anger vs. sadness, and joy vs. sadness. We thus concluded that F0–related parameters, voiced energy range, and the rate of delivery mainly contribute to the differentiation of emotions at the level of physiological arousal.

In order to find vocal indicators of emotional valence, we compared voice quality parameters for anger (a negative emotion with high level of physiological arousal) with those for joy (a positive emotion with high level of physiological arousal). This was inspired by the studies on the measurement of vocal differentiation of hedonic valence in spectral analyses of the voices of astronauts (Popov *et al.*, 1971; Simonov *et al.*, 1980). We thus hypothesised that spectral parameters could significantly differentiate between positive and negative valence of the emotions which have similar levels of physiological activation. To this purpose, long-term average spectra (LTAS) were computed for each voice sample, yielding 128 data points for a range of 40–5 500 Hz.

We used a Bark-based strategy of spectral data analyses, where perceptually equal intervals of pitch are represented as equal distances on the scale. The frequencies covered by 1.5 Bark intervals were the following: 40–161 Hz; 161–297 Hz;

Table 23.2 Partial correlation coefficients between mean F0 and other vocal parameters with speaker's gender partailed out

Mean F0 and emotions	F0 max/min ratio	F0 sd	voiced energy range in pseud dB	Delivery rate
mean F0 in Anger	.43**	.77**	−.03	.39**
mean F0 in Joy	.36**	.66**	−.08	.16
mean F0 in Sadness	.32**	.56**	−.43**	−.13

Note: N = 66. $*p < .05$, $**p < .01$, $***p < .001$; all significance levels are 2-tailed.

Table 23.3 Acoustic differentiation of emotions in male speakers

Emotions compared	F0 mean in Hz	T-test and P	F0 max/min ratio	T-test and P	F0 SD	T-test and P	Voiced energy range in pseudo d	T-test and P	Delivery rate	T-test and P
sadness	104		1.6		10.2		9.6		3.9	
anger	128	−4.3***	2.0	−6.0***	21.2	−5.7***	14.2	−5.0***	4.6	−2.2*
sadness	104		1.6		10.2		9.6		3.9	
joy	126	−4.6***	1.9	−6.0***	22.7	−7.5***	12.1	−2.5*	4.5	−2.9**
joy	126		1.9		22.7		12.0		4.5	
anger	128	−.4	2.0	−.9	21.2	.8	14.2	−2.8**	4.6	−.2

Note: N = 30. *$p < .05$, **$p < .01$, ***$p < .001$; all significance levels are 2-tailed.

Table 23.4 Acoustic differentiation of emotions in female speakers

Emotions compared	F0 mean in Hz	T-test and P	F0 max/min ratio	T-test and P	F0 SD	T-test and P	voiced energy range in pseudo dB	T-test and P	Delivery rate	T-test and P
Sadness	201		1.5		19.0		10.9		4.2	
Anger	228	−2.7**	1.8	−3.4**	33.8	−4.8***	14.2	−2.9**	5.0	−3.7**
Sadness	201		1.5		19.0		10.9		4.2	
Joy	236	−3.7**	1.9	−5.7***	37.0	−6.1***	12.8	−2.2*	5.0	−3.3**
Joy	236		1.9		37.0		12.8		5.0	
Anger	228	.8	1.8	1.6	33.8	1.0	14.2	−1.0	5.0	−.1

Note: N = 36. *$p < .05$, **$p < .01$, ***$p < .001$; all significance levels are 2-tailed.

297–453 Hz; 453–631 Hz; 631–838 Hz; 838–1 081 Hz; 1 081–1 370 Hz; 1 370–1720 Hz; 1 720–2 152 Hz; 2 152–2 700 Hz; 2 700–3 400 Hz; 3 400–4 370 Hz; 4 370–5 500 Hz (Hassal and Zaveri, 1979; Pittam and Gallois, 1986; Pittam, 1987). Subsequently mean energy value for each band was computed. We thus obtained 13 spectral energy values per emotion and per subject.

Paired t-tests were applied. The pairs consisted of the same acoustic parameter (the value regarding the same frequency interval) compared across two emotions. The results showed that several frequency bands contributed significantly to the differentiation between anger and joy, thus confirming the hypothesis that the valence dimension of emotions can be reflected in the long term average spectrum.

The results show that in a large portion of the spectrum, energy is higher in anger than in joy. In male subjects it is significantly higher as of 300 Hz up to 3 400 Hz, while in female subjects the spectral energy is higher in anger than in joy in the frequency range from 800–3 400 Hz. Thus our analysis of LTAS curves, based on 1.5 Bark intervals, shows that an overall difference in energy is not the consequence of major differences in the distribution of energy across the spectrum for Anger and Joy. This fact may lend itself to two interpretations: (1) those aspects of voice quality which are measured by spectral distribution are not relevant for the distinction between positive and negative valence of high-arousal emotions or (2) anger and joy also differ on the level of arousal which is reflected in spectral energy (both voiced and voiceless). Table 23.5 presents the details of the results for the Bark-based strategy of the LTAS analysis.

Although we assumed that vocal signalling of emotion can function independently of the semantic and affective information inherent to the text (Banse and Scherer, 1996; Scherer, Ladd, and Silverman, 1984), the generally positive connotations of

Table 23.5 Spectral differentiation between anger and joy utterances in 1.5 Bark frequency intervals.

Frequency bands in Hz	spectral energy in pseudo dB Male subjects	T-test and P	spectral energy in pseudo dB Female subjects	T-test and P
40–161	A 18.6; J 17.6	.69	A 12.2; J 13.8	−1.2
161–297	A 23.5; J 20.8	2.0	A 19.1; J 18.9	.12
297–453	A 26.7; J 22	3.1*	A 21.9; J 20.8	.62
453–631	A 30.9; J 24.3	3.4**	A 24.2; J 21.3	1.5
631–838	A 28.5; J 21.0	4.4**	A 23.6; J 19.3	2.2
838–1 081	A 21.1; J 15.8	3.8**	A 19.4; J 14.7	2.6*
1 081–1 370	A 19.6; J 14.8	3.6**	A 16.9; J 12.6	2.9*
1 370–1 720	A 22.5; J 17.0	3.7**	A 17.5; J 12.9	3.3**
1 720–2 152	A 20.7; J 14.6	3.8**	A 19.7; J 16.1	2.5*
2 152–2 700	A 18.7; J 13.0	3.7**	A 15.2; J 12.4	2.4*
2 700–3 400	A 13.3; J 10.1	2.9*	A 14.7; J 11.3	2.7*
3 400–4 370	A 10.6; J 4.1	2.5	A 8.8; J 3.9	1.7
4 370–5 500	A 1.9; J .60	1.2	A 1.3; J .5	1.9

Note: N = 20 *p < .05, **p < .01, ***p < .001; A = anger; J = joy; All significance levels are 2-tailed.

the words 'accept' and 'deal' sometimes did disturb the subjects' ease of saying the sentence with a tone of anger. Such cases were not taken into account for statistical analyses. However, this fact points to the influence of the semantic content on vocal emotional expression. Most of the subjects reported that emotionally congruent semantic content could considerably help produce appropriate tone of voice. The authors also repeatedly noticed that in the subjects(spontaneous verbal expression, the emotion words were usually said on an emotionally congruent tone.

Conclusion

In spite of remarkable individual differences in vocal tract configurations, it appears that vocal expression of emotions exhibits similar patterning of vocal parameters. The similarities may be partly due to the physiological factors and partly to the contextually driven vocal adaptations governed by stereotypical representations of emotional voice patterns. Future research in this domain may further clarify the influence of cultural and socio-linguistic factors on intra-subject patterning of vocal parameters.

Acknowledgements

The authors thank Jacques Terken, Technische Universiteit Eindhoven, Nederland, for his constructive critical remarks. This article was carried out in the framework of COST 258.

References

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

Hassal, J.H. and Zaveri, K. (1979). *Acoustic Noise Measurements*. Büel and Kjaer.

Keller, E. (1994). *Signal Analysis for Speech and Sound*. InfoSignal.

Mendolia, M. and Kleck, R.E. (1993). Effects of talking about a stressful event on arousal: Does what we talk about make a difference? *Journal of Personality and Social Psychology*, *64*, 283–292.

Pittam, J. (1987). Discrimination of five voice qualities and prediction of perceptual ratings. *Phonetica*, *44*, 38–49.

Pittam, J. and Gallois C. (1986). Predicting impressions of speakers from voice quality acoustic and perceptual measures. *Journal of Language and Social Psychology*, *5*, 233–247.

Popov, V.A., Simonov, P.V. Frolov, M.V. *et al.* (1971). Frequency spectrum of speech as a criterion of the degree and nature of emotional stress. (Dept. of Commerce, JPRS 52698.) *Zh. Vyssh. Nerv. Dieat.*, (*Journal of Higher Nervons Activity*) *1*, 104–109.

Scherer, K.R. (1981). Vocal indicators of stress. In J. Darby (ed.), *Speech Evaluation in Psychiatry* (pp. 171–187). Grune and Stratton.

Scherer, K.R. (1989). Vocal correlates of emotional arousal and affective disturbance. *Handbook of Social Psychophysiology* (pp. 165–197). Wiley.

Scherer, K.R. (1992). On social representations of emotional experience: Stereotypes, prototypes, or archetypes? In M.V.H Cranach, W. Doise, and G. Mugny (eds), *Social Representations and the Social Bases of Knowledge* (pp. 30–36). Huber.

Scherer, K.R. (1993). Neuroscience projections to current debates in emotion psychology. *Cognition and Emotion*, *7*, 1–41.
Scherer, K.R. and Kappas, A. (1988). Primate vocal expression of affective state. In D.Todt, P.Goedeking, and D. Symmes (eds), *Primate Vocal Communication* (pp. 171–194). Springer-Verlag.
Scherer, K.R., Ladd, D.R., and Silverman, K.E.A. (1984). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America*, *76*, 1346–1356.
Scherer, K.R. and Zei, B. (1988). Vocal indicators of affective disorders. *Psychotherapy and Psychosomatics*, *49*, 179–186.
Simonov, P.V., Frolov, M.V., and Ivanov E.A. (1980). Psychophysiological monitoring of operator's emotional stress in aviation and astronautics. *Aviation, Space, and Environmental Medicine*, January 1980, 46–49.
Williams, C.E. and Stevens, K.N. (1972). Emotion and speech: Some acoustical correlates. *Journal of the Acoustical Society of America*, *52*, 1238–1250.
Williams, C.E. and Stevens, K.N. (1981). Vocal correlates of emotional states. In J.K. Darby (ed.), *Speech Evaluation in Psychiatry* (pp. 221–240). Grune and Statton.
Zei, B. and Archinard, M. (1998). La variabilité du rythme cardiaque et la différentiation prosodique des émotions, *Actes des XXIIèmes Journées d'Etudes sur la Parole* (pp. 167–170). Martigny.