EXTRACTION OF EMOTIONAL PROSODY FROM TELEPHONE CALLS TO HOSPITAL EMERGENCY DEPARTMENTS

Dongjun Wei[1], Mohamed Embarki[1], Oussama Barakat[2], Coralie Vaucherey[1,2], Thibaut Desmettres[2,3], Tania Marx[2,4] et Stephan Robert[5]

[1]ELLIADD EA 4661, université de Franche-Comté, Besançon

[2]Laboratoire de Nanomédecine, Imagerie et Thérapeutique EA 4662, université de Franche-Comté, Besançon

[3]Hopitaux universitaires Genève, université de Genève [4]Centre Hospitalier Régional Universitaire de Besançon

[5]Haute Ecole d'Ingénierie et de Gestion du Canton de Vaud, Yverdon-les-Bains

dongjun.wei@edu.univ-fcomte.fr, mohamed.embarki@univ-fcomte.fr, oussama.barakat@univ-fcomte.fr, coralie.vaucherey@edu.univ-fcomte.fr, thibaut.desmettre@univ-fcomte.fr, tania.marx@univ-fcomte.fr, stephan.robert@heig-vd.ch

ABSTRACT

This study is based on the exploitation of telephone calls to hospital emergency departments in order to train artificial intelligence tools for the automatic recognition of vocal emotions for the purpose of improving emergency medical regulation. The main objective is to extract the most characteristic acoustic cues of vocal emotions. Very limited samples were selected from the audio databases of two University Hospitals in the Jura Arc (CHRU of Besançon in France and CHUV of Lausanne in Switzerland). Only extracts containing linguistic vocal emotions with a negative polarity (angst, anger, embarrassment and sadness) were retained for the study. These extracts were segmented and labeled, and their acoustic parameters (*F0*, intensity and duration) were measured manually. The preliminary results show that *F0* modulations seem more accurate to discriminate the four vocal emotions.

Keywords: vocal emotion, *F0*, emergency medical regulation, artificial intelligence

1. INTRODUCTION

Automatic Speech Recognition (ASR) has made significant progress both in terms of approaches and models [1], speaker recognition has also reached important steps [2]. Speech conveys both linguistic and non-linguistic information, such as emotions. The human ability to recognize emotions from verbal vocalizations has been repeatedly demonstrated [3]. Also, from non-linguistic vocalizations listeners are able to identify different types of emotion [4]. Apart from the lexicon, which can easily be attached to a type of emotion, prosody has been described as an important source in the expression of emotions [5]. Based on prosodic features, listeners have the ability to identify different types of emotions not only within their culture, but also cross-culturally [6,7]. Automatic emotion recognition has been developed for various applications [8]. A growing need for vocal emotion recognition systems is needed in the field of emergency medical regulation in order to be able to quickly identify the seriousness of the situation, direct towards the right course of care and/or engage the appropriate means of transport.

2. VOCAL EMOTIONS

Emotions cannot be reduced to a single process but are part of a complex set of processes [9]. They are grouped into two categories, basic emotions and extended emotions [10, 11]. Both are expressed by the modulation of linguistic (verbal) and/or non-linguistic (laughing, crying, screaming, moaning) vocalizations [12]. Vocal emotion research has focused more on basic emotions and their acoustic properties which are of 3 types: prosodic, spectral and voice quality. During strong emotional states, such as anger, significant acoustic modulations are visible, speech becomes higher pitched, shorter and more intense in auditory terms, faster and with less pauses. During a moderate emotional state, such as sadness, the speaker's voice becomes deeper, less intense, and the tone deeper, the rate slower, with more pauses. Ultimately, any modification of the emotional state is conveyed by modulations of the acoustic properties, such as the fundamental frequency (*F0*), the intensity and the duration [13]. Among all these cues, *F0* is generally recognized to discriminate more accurately vocal emotions.

In order to design an automatic vocal emotion recognition system suitable for the regulation of medical emergencies, it is necessary to better understand their acoustic properties.

3. METHODS

The audio data used in this study come from telephone calls to the medical emergency services of two hospitals, the CHRU of Besançon (France) and the CHUV of Lausanne (Switzerland). The telephone calls used are balanced in number (16 calls from each hospital) and gender (16 women and 16 men). The calls were first anonymized before processing. Two native French evaluators were recruited to label negative valence vocal emotions (angst, anger, embarrassment, fear and sadness). The labeling of vocal emotions was largely based on the subjective feelings of the labelers, however a defining explanation had been provided to them in order to better distinguish the callers emotions:

1. angst: anticipation of future threats
2. anger: reaction to the fact of not being respected, of not having one's needs met.
3. Embarrassment: reactions when one's behavior deviates from and falls below normative social standards.
4. fear: reaction to a real or imminent threat in the present.
5. sadness: reaction following the experience of an irreparable loss.

For each labeled extract, the unanimity of the two evaluators was required. In case of disagreement between

the two, a third evaluator was asked and if the disagreement persisted, the extract was removed from the sample. The extracts were labeled in Praat [14] and at the end of this step, it was observed that the evaluators frequently confused angst with fear. The choice was made to group the two vocal emotions within the angst category. The acoustic properties of the extracts selected in this study were measured manually using Praat. Non-linguistic vocal emotions were excluded from the study, and the linguistic vocal emotions retained all correspond to complete utterances, *i.e.* a total of 406 utterances.

The *F0* measurements targeted the four inflection points of each utterance (onset, offset, maximum and minimum) as well as the *F0* average. The conversion formula is 12*Log2(*F0* target/*F0* reference)=tonal difference (semitone) and 12 semitone=1 octave, which is converted to a chromatic gap when the *F0* gap is involved. The same points were measured for the intensity. Duration measurements targeted articulation rate and speech rate.

4. RESULTS

The abbreviations used in all tables are listed first by category as follows:

1) Ax.= angst, Ag=anger, E.=embarrassment, S.=sadness
2) Max.= maximum, Min.= minimum, Ons.=Onset, Off.=Offset, Avg.=average
3) W.=woman, M.=man
4) Fr.=France, Sw.=Switzerland
5) *F0*A=*F0* average, AI=intensity average, ARA= articulation rate average, ASR= speech rate average
6) SD=standard deviation, ST=semitone, ID=intensity deviation, DD=duration deviation

406 linguistic vocal emotions were classified according to the type of emotion (angst, anger, embarrassment and sadness), gender of the patient and his/her country：274 French extracts and 132 Swiss extracts were analyzed. Table 1 shows a wide variation in the recognition of the 4 emotions by the evaluators: with 84% for angst and 11%, 4% and 1% for embarrassment, sadness and anger respectively. Angst vocal emotion ~~has~~ obtained the highest recognition rate, while embarrassment, sadness and anger ~~have~~ had a lower recognition rate. This suggests that although embarrassment, sadness and anger vocal emotions are serious negative emotions, they are less common due to their occurrence in phone calls to medical emergency departments.

	Ax.		Ag.		E.		S.		
	M	W	M.	W	M	W	M	W	
All	342		5		43		16		406
	197	145	2	3	38	5	13	3	
Fr.	214		5		41		14		274
	127	87	2	3	36	5	12	2	
Sw.	128		0		2		2		132
	70	58	0	0	2	0	1	1	

Table 1: Number of linguistic vocal emotions classified by type of emotion, gender and country.

The following tables present acoustic measurements. Tables 2, 3 and 4 relate~~s~~ to *F0* measurements; Tables 5, 6 and 7 to intensity measurements and Tables 8 and 9 to duration measurements. Cells in the tables without data are marked in grey.

Table 2 deals with the *F0*A (Hz) and semitone difference (ST) of the four labelled emotions. ST 1 (2.3.4) represents gendered tonal difference in the same country, while ST5 represents tonal difference between patients of the same gender across countries.

		Fr.		Sw.		ST5
		*F0*A	SD	*F0*A	SD	
Ax.	W.	231	32	232	30	0
	M.	151	20	142	24	1
	ST1	7		8		
Ag.	W.	226	22			
	M.	143	10			
	ST2	8				
E.	W.	224	53			
	M.	133	19	108	19	4
	ST3	10				
S.	W.	232	40	331		6
	M.	131	16	101		5
	ST4	10		20		

Table 2: *F0*A (Hz), SD and ST in French and Swiss patients in four vocal emotions.

Table 2 reflects the following general patterns: 1. Women in different countries had higher *F0* average in angst and sadness vocal emotions and lower ones in embarrassment and anger, whereas men had higher ones in angst and anger vocal emotions and lower ones in embarrassment and sadness. 2. As far as angst vocal emotion is concerned, *F0* average and its differences were narrower across countries.

Table 3 deals with *F0* average (Hz) and tonal difference (ST) for inflection points for different callers in angst and anger vocal emotions. ST1 (2, 3 and 4) indicates tonal difference of the same gender across countries. ST5 (6) indicates gendered tonal difference of the same country. ST7 (8) relates to the tonal difference between *F0* onset/offset and *F0* maximum/minimum of the same gender in the same country.

		Ax.			Ag.		
		W.	M.	ST5	W.	M.	ST6
Ons.	Fr.	224	142	8	228	142	8
	Sw.	234	138	9			
	ST1	1	1				
Off.	Fr.	226	159	6	181	120	7
	Sw.	224	141	8			
	ST2	0	2				
ST7	Fr.	0	2		4	3	
	Sw.	8	0				
Max.	Fr.	360	237	7	345	220	8
	Sw.	323	209	8			
	ST3	2	2				
Min.	Fr.	137	108	4	155	73	13
	Sw.	162	99	9			
	ST4	2	1				
ST8	Fr.	17	14		14	19	
	Sw.	12	13				

Table 3: *F0*A (Hz) and ST at four inflection points in French and Swiss callers in angst and anger vocal emotions.

Table 3 shows the following general patterns: In angst vocal emotion, *F0* onset and offset average were similar for French women, whereas *F0* onset average was higher than the offset for Swiss women (8 semitones). French men *F0* onset average was lower than the offset (2 semitones); Swiss men *F0* onset and offset average was similar. Differences between *F0* maximum and minimum average

for men and women in both countries were not less than 1 octave.

Table 4 deals with *F0* average (Hz) with tonal difference (ST) at four inflection points in embarrassment and sadness vocal emotions.

		E.			S.		
		W.	M.	ID5	W.	M.	ID6
Ons.	Fr.	220	133	9	263	138	11
	Sw.		103		305	106	18
	ST1		4		3	6	
Off.	Fr.	205	138	7	217	111	12
	Sw.		118		447	89	28
	ST2		3		13	4	
ST7	Fr.	1	1		3	4	
	Sw.		2		7	3	
Max.	Fr.	328	217	7	291	194	7
	Sw.		142		463	115	24
	ST3		7		8	9	
Min.	Fr.	142	93	7	132	94	6
	Sw.		88		254	80	20
	ST4		1		11	3	
ST8	Fr.	15	15		14	13	
	Sw.		8		10	6	

Table 4: *F0*A (Hz) and ST at four inflection points in French and Swiss patients in embarrassment and sadness vocal emotions.

Table 4 reflects the following general pattern: 1. In ~~the~~ embarrassment vocal emotion, women's *F0* onset average in both countries was higher than the offset, 1 and 8 semitones for French and Swiss women respectively. In both countries, men's *F0* onset average was lower than ~~at~~ the offset (no less than 1 semitone). 2. In both countries, *F0* onset average for sadness vocal emotion was lower than the offset for men and women (not less than 3 semitones). The difference between *F0* maximum and minimum average for both French men and women was more than 1 octave~~,~~ while it was within 1 octave for both Swiss men and women.

Table 5 presents the intensity average (dB) and intensity deviation (ID) in different countries in the four vocal emotions. ID1 (2.3.4) represents the intensity average differences across genders in the same country, while ID5 represents the difference for the same gender across countries.

		Fr.		Sw.		ID5
		AI	SD	AI	SD	
Ax.	W.	74	3	75	7	1
	M.	71	5	72	5	1
	ID1	3		3		
Ag.	W.	74	3			
	M.	73	4			
	ID2	1				
E.	W.	71	4			
	M.	70	6	73	2	3
	ID 3	1				
S.	W.	74	3	80		6
	M.	74	2	69		5
	ID 4	0		11		

Table 5: intensity average (dB), standard deviation and intensity deviation (ID) for French and Swiss patients in the four vocal emotions.

The following general pattern is reflected in Table 5: intensity average in the four vocal emotions were very close. Table 6 deals with intensity average and intensity deviation (ID) at four inflection points in angst and anger vocal emotions. ID7 (8) relates to the difference in intensity deviation between the onset/offset and the maximum/minimum for the same sex in the same country.

		Ax.			Ag.		
		W.	M.	ID5	W.	M.	ID6
Ons.	Fr.	67	67	0	65	66	1
	Sw.	71	67	3			
	ID1	4	0				
Off.	Fr.	62	62	0	64	63	1
	Sw.	64	61	3			
	ID2	2	1				
ID7	Fr.	5	5				
	Sw.	7	6				
Max.	Fr.	82	82	0	87	81	6
	Sw.	84	81	3			
	ID3	2	1				
Min.	Fr.	28	28	0	25	32	7
	Sw.	35	31	4			
	ID4	7	7				
ID8	Fr.	54	54				
	Sw.	49	50				

Table 6: intensity average (dB) and intensity deviation (ID) at four inflection points in French and Swiss patients in angst and anger vocal emotions.

Table 6 shows the following general pattern: for angst and anger vocal emotions, onset and offset intensity average were not very different from patient to patient, as was intensity average at the maximum and minimum.

Table 7 deals with intensity average (dB) and intensity deviation (ID) at the four inflection points in embarrassment and sadness vocal emotions.

		E.			S.		
		W.	M.	ID5	W.	M.	ID6
Ons.	Fr.	67	68	1	70	69	1
	Sw.		71		72	75	3
	ID1		3		2	6	
Off.	Fr.	55	61	6	75	64	11
	Sw.		62		62	56	6
	ID2		1		13	8	
ID7	Fr.		7		5	5	
	Sw.		8		10	9	
Max.	Fr.	80	80	0	82	83	1
	Sw.		82		88	78	11
	ID3		2		6	5	
Min.	Fr.	26	27	1	72	29	43
	Sw.		41		46	26	20
	ID4		13		26	3	
ID8	Fr.		53		10	54	
	Sw.		41		42	52	

Table 7: intensity average (dB) and intensity deviation (ID) at four inflection points in French and Swiss patients with embarrassment and sadness vocal emotions.

The following general pattern is reflected in Table 7: as for the two previous vocal emotions, onset and offset intensity average in embarrassment and sadness were not very pronounced, and the same was true for intensity average at the maximum and the minimum.

Tables 8 and 9 concern the speech rate and articulation rate in the two countries in each of the four vocal emotions. DD1 (2.3.4) indicates the duration difference in the same gender in different countries, whereas DD5 (6) indicates gendered duration difference in the same country.

	Fr.				DD5
	SRA	SD	ARA	SD	

Ax.	W.	5	1	5	1	0
	M.	5	2	5	1	0
	DD1	0		0		
Ag.	W.	5	0	6	0	1
	M.	3	0	4	0	1
	DD2	2		2		
E.	W.	5	1	6	1	1
	M.	5	1	5	1	0
	DD3	0		1		
S.	W.	5	2	5	1	0
	M.	4	2	5	2	1
	DD4	1		0		

Table 8: duration average (syl/s), standard deviation and duration difference (syl/s) in French patients in four vocal emotions.

		Sw.				DD5
		SAR	SD	ARA	SD	
Ax.	W.	5	1	5	1	0
	M.	5	2	6	2	1
	DD1	0		1		
Ag.	W.					
	M.					
	DD2					
E.	W.					
	M.	4	2	4	2	0
	DD3					
S.	W.	4		4		0
	M.	4		5		0
	DD4	0		1		

Table 9: duration average (syl/s), standard deviation and duration difference (syl/s) in Swiss patients in four vocal emotions.

The following general pattern is reflected in Tables 8 and 9: the differences in speech rate and articulation rate between callers were not very significant in the four vocal emotions.

5. DISCUSSION

Regardless of gender or country, data presented here showed that linguistic vocal emotions were discriminated mainly by *F0* modulations, intensity and duration showed minor distinctions. *F0* differences of vocal emotions can be broadly summarized as follows:

5.1 *F0* average

5.1.1 Same gender in the same country

French women highest vocal emotion *F0* average was, in descending order, sadness, angst, anger and embarrassment, whereas French men highest *F0* average was anxiety, followed by anger, embarrassment and finally sadness. The highest *F0* average among Swiss women was for sadness, followed by anxiety (as a reminder, data were missing for anger and embarrassment). Among Swiss men, the highest F0 average was for anxiety, followed by embarrassment and finally sadness (data were missing for anger).

5.1.2 Gender differences in the same country

F0 average was consistently higher for French women than for men across the four vocal emotions, with the largest tonal difference in sadness and embarrassment and the smallest in angst and anger. For Swiss women, *F0* average was consistently higher than for men in all four vocal emotions, with the largest tonal difference in sadness, which remained above 1 octave, and the smallest in angst, which remained within 1 octave.

5.1.3 Same gender across countries

F0 average was higher for Swiss women than for French in sadness, with tonal difference within 1 octave, while in angst, the difference was not significant. *F0* average was higher for French men than for Swiss in angst, embarrassment and sadness, but the tonal difference remained within 1 octave.

5.2. *F0* average at four inflection points

French women *F0* onset average was generally higher in anger, embarrassment and sadness vocal emotions than the offset, with a tonal difference of 1 octave or less, while the difference was not significant in angst vocal emotion. In contrast, *F0* maximum average was 1 to 2 octaves higher than *F0* minimum average. French men *F0* onset average in angst and embarrassment vocal emotions was generally 3 semitones lower than the offset, whereas in anger and sadness vocal emotions the onset was 4 semitones higher than the offset. Swiss men *F0* onset average was generally 2 semitones lower than the offset in angst and embarrassment vocal emotions whereas it was 3 semitones higher than the offset in sadness.

5.2.1 Gender difference in the same country

French women *F0* onset and offset average in the four vocal emotions was consistently higher than for men, however tonal difference remained within 1 octave for all of them. The same was true for *F0* maximum and minimum average. Swiss women onset and offset *F0* average in angst and sadness vocal emotions were consistently higher than those of men, with a tonal difference of approximately 1 octave or less in angst and more than 1 octave in sadness. The same was true for *F0* maximum and minimum average.

5.2.2 Same gender across countries

French women *F0* onset average in angst and sadness vocal emotions was 3 semitones lower than that of Swiss women –It was the opposite for *F0* offset average, with a small tonal difference in angst vocal emotion and a tonal difference of more than 1 octave in sadness. French women *F0* maximum and minimum average was lower than that of Swiss women, tonal difference was within 1 octave. The situation was reversed for men in both countries, tonal difference was also within 1 octave.

Thus, vocal emotion acoustic properties in both countries showed different *F0* patterns, and vary very little in intensity and duration. Swiss Results were more complex and lack sufficient data support for some vocal emotions (anger and embarrassment), either because the prevalence of the Swiss accent compared to the French accent influences the labelers' judgment, the relatively small number of overall emotion samples, or because the corpus is composed of purely natural speech, which differs markedly from previous studies based on a large corpus of simulated emotions.

6. CONCLUSION

The results show that, among the 3 acoustic properties analyzed which are *F0*, intensity and duration, the modulations of *F0* are the most representative of the change in linguistic vocal emotion, and more particularly the modulations in *F0* contour and *F0* average. On the other hand, the intensity and the duration distinguish only very weakly the linguistic vocal emotions. Therefore, in the design of automatic voice emotion recognition system, we can assume that the data extraction of *F0* remains the basic acoustic property for natural linguistic voice emotion recognition.

7. REFERENCES

[1] Ainsworth, W.A. 1999. Some approaches to automatic speech recognition. In: Hardcastle, W., Laver, J. (eds), *The Handbook of Phonetic Sciences*. Blackwell, 721-743.
[2] Nolan, F. 1999. Speaker recognition and Forensic phonetics. *The Handbook of Phonetic Sciences*. Blackwell, 744-767.
[3] Juslin, P.N., Laukka, P. 2003. Communication of Emotions in Vocal Expression and Music PerFormance: Different Channels, Same Code? *Psychological Bulletin* 129, 770-814.
[4] Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. 2016. The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion* 16, 117–128.
[5] Keltner, D., Tracy, J., Sauter, D. A., Cordaro, D. C., & McNeil, G. 2016. Expression of emotion. In L. F. Barrett & M. Lewis (eds.), *Handbook of emotions*, GuilFord Press, 467-482.
[6] Cowen, A.S., Elfenbein H.A. Laukka, P. & Keltner, D. 2019. Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist* 74, 698-712.
[7] Cowen, A.S., Laukka, P., Elfenbein H.A., Liu, R. & Keltner, D. 2019. The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature Human Behaviour* 3, 369-382.
[8] Yanga, N., Deyb, N., Sherrattc, R.S. & Shia, F. 2020. Recognize basic emotional statesin speech by machine learning techniques using mel-frequency cepstral coefficient features. *Journal of Intelligent & Fuzzy Systems* 39, 1925-1936.
[9] Wager, Tor D., Krishnan A., Hitchcock, E. 2018. How are emotions organized in the brain? *The nature of emotion: Fundamental questions*. OxFord, 112-118.
[10] Ekman, P. 1972. Universal and cultural differences in facial expression of emotions. In Cole, J. (ed.), *Nebraska symposium on motivation*. Lincoln; 207-283.
[11] Plutchik, R., Kellerman, H. (eds.). 1980. Emotion: theory, research, and experience. *Theories of emotion*. Academic Press, Vol. 1, 3-33.
[12] Bryant, G.A. 2021. The evolution of human vocal evolution. *Emotion Review* 13, 25-33.
[13] Scherer, K. R. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40, 227–256.
[14] Boersma, P. (2017). Praat, a system For doing phonetics by computer. *Glot International* 5:9/10, 341-345.