Perceptual and acoustic correlates of spontaneous vs. social laughter

Takaaki Shochi[1],[2], *Marine Guerry*[1], *Jean-luc Rouas*[2], *Marie Chaumont*[3], *Toyoaki Nishida*[4], *Yoshimasa Ohmoto*[4]

[1]LaBRI, Univ. Bordeaux, CNRS UMR 5800, 33400 Talence, France
[2]CLLE-ERRSaB UMR5263, Bordeaux, France
[3]Bordeaux Montaigne University, France
[4]Kyoto University, Japan

takaaki.shochi@labri.fr, marine.guerry@etu.u-bordeaux-montaigne.fr, jean-luc.rouas@labri.fr, marie.chaumont@etu.u-bordeaux-montaigne.fr, nishida@i.kyotou.-ac.jp, ohmoto@i.kyoto-u.ac.jp

Abstract

The current paper focuses on the various types of laughter recorded during real social interactions in a virtual immersive environment. In this experiment, we investigate whether human beings are able to discriminate perceptually determined volitional social laughs from spontaneous involuntary laughs using only audio information without any contextual cues. Towards this aim, we designed a perceptual experiment taken by 82 French and 20 Japanese subjects. Each subject listened to 162 laughs and chose one response among three possibilities : social, spontaneous or unknown (I don't know). The results show that all listeners are able to discriminate these two types of laughter with quite good confidence without contextual information : the correct identification rate for spontaneous laughter is about 70% with a similar amount for social laughter. We then extracted acoustic characteristics for each laughter in order to investigate possible differences between the two types of laughter. Moreover, multiple factor analysis shows that perceptual behaviours and some acoustic features (F0 and duration) are correlated. Especially, we observe a significant difference between social and spontaneous laughter through the features of F0 and total duration.

Index Terms : human laughter, recognition, acoustics, perception, culture

1. Introduction

Comprehensive knowledge about the vocal characteristics of social affective interaction has been neglected for a long time because of the lack of sufficient understanding about cognitive processing of various affective meanings as well as technical realization of such expressions. However, automatic recognition and synthetic realization of such affective meaning became one of the important issues for researchers of various scientific research fields like social robotics, medical hearing tools or language learning tools etc. [1, 2].

Such human social interaction is an exchange of social information conveyed by voice, eye contact, gestures, facial expressions, sighs or laughter [2, 3, 4]. Among these modalities, laughter must be one of the most important behaviours in the development of speech and in human and animal communication [5, 6, 7]. Laughter is often considered as a physical reaction to external stimuli which are often linked to positive valence (i.e. joyful reaction). Although laughter is deeply rooted in human biology, it also serves very strong social roles to bring about positive, mutually beneficial relationships among people and communities [8, 9]. In [10], the authors reported that laughter is usually provoked by external stimuli, and organized on three different axes : neuro-hormonal involving periaqueductal gray, the reticular formation with inputs from cortex basal ganglia and the hypothalamus [11], including muscular inputs and the respiratory axis.

In [9], the authors suggest the existence of two different types of laughter : spontaneous and volitional (or social) by neuro-physiological differences. Spontaneous laughter is considered an involuntary reaction to external stimuli. It is supposed to be innate because it occurs even before the first words. Physiological changes during such involuntary laughter are quite different from what occurs during a voluntary one. For instance, involuntary laughter is characterized by a higher activation of hypothalamus than for the voluntary one, and the chest expansion and amplitude of sound waves show more regular cycle patterns than the voluntary one which exhibits a speech-like pattern. On the other hand, social laughter is supposed to be an intentional communicative act in order to set up a positive relationship or to tone down the conflictive tension.

Concerning the acoustic realization of these various types of laughter, [12] suggests three levels of description : "bouts", "calls" and "segments". With regard to segmentation, [13] made a distinction between "spontaneous" and "social" laughter. According to recent work ([14], [12], [15], [16]), the spontaneous is higher and has a more variable F0, as well as higher variability in acoustic parameters in general. In addition, the spontaneous laughter is also characterized by longer duration with a shorter burst duration, ingressive and chuckle sounds ([13], [17]). However, there is no significant difference for both types of laughter regarding the breathiness and the mouth aperture. According to our assumption, (1) human beings are able to discriminate between social (voluntary) laughs and spontaneous (involuntary) ones using only audio information without any context. (2) Perceptually determined spontaneous laughter may have common acoustic cues among different cultures. On the contrary, (3) volitional social laughter may be perceived differently from one culture to another following cultural conventional manners.

Following these hypotheses, the current research investigates (1) whether French and Japanese subjects can discriminate between social volitional laughter and spontaneous involuntary laughter using only auditory laughs extracted from an immersive virtual interaction without any context. Independently we aim at investigating (2) acoustic characteristics for each type of laughter in French and Japanese.

2. Corpus

The stimuli were recorded in an immersive virtual environment at Kyoto University, Japan. This database consists of spontaneous affective speech recorded during a virtual reality game played by three participants. The game was designed to study communications between people in virtual environments and was made using Unity. Each player was alone in his own individual immersive virtual environment (completely surrounded by displays or in an immersive dome), but could communicate with the others using cameras and microphones. They were required to communicate in order to solve various tasks instructed by three different virtual characters. One of the main interests of this approach is that each participant can be recorded individually. A total of 12 spontaneous affective speech data files 9 Japanese (2F/7M) and 3 French (1F/2M) were recorded. A total of 254 sequences containing only laughter were manually segmented using PRAAT [18]. A first pilot test was conducted in order to investigate what acoustic features distinguish spontaneous emotional vocalizations of laughs from volitional forms which are considered as social laughs ([17], [14], [19]). 7 experimenters (3 Japanese males and 4 French (3F/1M) are instructed to annotate each sample using two labels spontaneous and social. According to a selection threshold criterion based on more than 70% of identification of the stimulus perceived as "spontaneous", a set of 27 spontaneous laughs was chosen. Another pilot test designed to choose the 27 volitional social laughs was done under the same criterion as for the spontaneous one by the 4 French experimenters who participated in the first pilot test.

3. Perceptual experiment

3.1. Paradigm

82 French native listeners (48F/34M, Mean age = 22.39 years) and 20 Japanese native listeners (9F/11M, Mean age = 24.55 years) were recruited in both countries. The stimuli were displayed 3 times each in audio alone condition in a randomized order (54 laughs (27 spontaneous / 27 social) x3 (repetitions) = 162 stimuli).

Before the test, subjects were informed about the definition of each type of laughter and the procedure of the experiment. The test was conducted individually using a GUI based interface developed under the "OpenSesame" software [20]. The total duration of the session took about 25 minutes. The subjects were required to listen to each stimulus at least once but could listen to the stimulus a second time maximum. Then, they had to select one choice among three possible answers : "spontaneous", "social", "I dont know". In the cases when spontaneous or social were selected, the subjects had to select a degree of certainty on a scale from 1 (not sure) to 7 (very sure). Definitions of the type of laughter provided in the instruction were :

— Spontaneous : it seems to you that the person is laughing in a spontaneous manner to an external event (e.g. a funny clip)

— Social : it seems to you that the person is laughing to maintain the communication with the other (e.g. embarrassed laughter, polite laughter, cynical laughter...)

3.2. Results

First of all, the χ^2 test was computed to investigate whether the distributions of listeners responses (Social, Spontaneous or Unknown) are independent or correlated. According to the result, a significant difference of the distribution of answers was observed (χ^2= 5284.7, ddl :2, p $<$ 0.001). According to the Table 1 (stimuli are in rows and the responses given by the subjects are in columns), the two types of laughter are well recognized : French subjects identified 69.24% for spontaneous laughs and 69.41% for social laughs ; Japanese listeners recognised 70.49% for spontaneous laughs and 74.63% for social laughs. These results confirmed that the listeners of both groups were able to recognize 2 types of laughter without visual indices or context.

TABLE 1: *Results for the perceptual test for French and Japanese listeners. Raw results are presented with their frequency for each row*

FRENCH	Spontaneous	Social	Unknown
Spontaneous	4599 (69.24%)	1683 (25.34%)	360 (5.42%)
Social	909 (13.69%)	4444 (69.41%)	1289 (19.41%)
Total result	5508 (41.46%)	6127 (46.12%)	1649 (12.41%)
JAPANESE	Spontaneous	Social	Unknown
Spontaneous	1142 (70.49%)	289 (17.84%)	189 (11.67%)
Social	370 (22.84%)	1209 (74.63%)	41 (2.53%)
Total result	1512 (46.67%)	1498 (46.23%)	230 (7.10%)

3.3. Correspondence analysis

In order to observe the perceptual distance of all responses based on the classification made by the listeners (spontaneous, social, I don't know) for 54 stimuli, we computed a Correspondence Analysis (CA) using FactoMineR package ([21]) under R software. According to the CA, the perceptual behaviour for 26 stimuli in the French group and 22 stimuli in the Japanese one, listeners showed an important contribution (i.e. above the expected average contribution for both 1st and 2nd dimensions).

Figure 1 and Figure 2 describe the distribution of 26 perceptual points for French and 22 for Japanese subjects on two psychometrical dimensions. The blue points on the figures represent the distribution of the perceptual behaviour and the three triangles represent the concept subjects have of the three answers. These two figures indicate that both French and Japanese listeners discriminate clearly the two types of laughter. It is also important to note that social and unknown categories are close together on the 1st dimension and far from spontaneous, which represents a well discriminated category. It indicates that volitional social laughs are more difficult to perceive than spontaneous ones. French subjects felt more difficulty to identify 5 laughs (located in the category of "unknown") rather than Japanese who had only two laughs in this category).

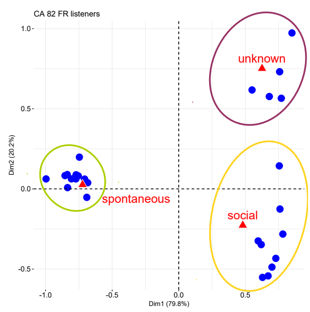

FIGURE 1: *Distribution of the perceptual behaviour of the French listeners for 26 stimuli*

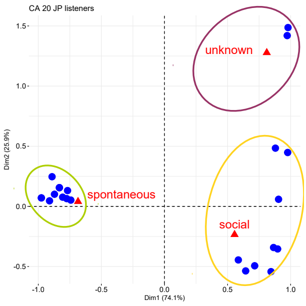

FIGURE 2: *Distribution of the perceptual behaviour of the Japanese listeners for 22 stimuli*

4. Acoustic analysis

For the purpose of the acoustic analysis, we measured several acoustic features that were previously reported to predict affective ratings and categorization for laughter as well as for more general affective voice analysis [17].

4.1. Features extraction

For the acoustic analysis, fundamental frequency (F0) and intensity are computed every 10 ms. They were extracted using a customized version of the Snack toolkit [22]. Most analyses are carried out on the voiced parts of the laughter as detected by the F0 extraction algorithm, thus ignoring non-voiced segments.

We extracted a set of 14 features in four main categories : F0 values for assessing the variability of the fundamental frequency (we expect, for instance, to have higher frequencies as well as more variability for spontaneous laughs), Intensity values - where higher levels and variability are also expected for spontaneous laughs, Duration values - social laughs are expected to be shorter and less voiced, Harmonics-to-noise ratios which were not explored in previous laughter studies but are expected to measure to some extent the breathiness level.

- F0mean (Hz) : the mean value of F0 extracted on voiced parts of the laughs
- F0SD : the standard deviation of F0 values on a laughter excerpt (voiced parts)
- F0slope (Hz/s) : the approximated slope of F0 (voiced parts only)
- NRJmean (dB) : mean of intensity values (whole file)
- NRJsd (dB) : standard deviation of intensity values during a laughter
- NRJslope (dB/s) : approximated slope of intensity during a laughter
- total.duration : duration of a manually annotated laughter
- voiced.duration : duration of all the voiced parts of a laughter
- NBvoiced : number of voiced segments
- HNR05 : harmonic to noise ratio in the frequency band between 0 and 0.5 kHz
- HNR15 : harmonic to noise ratio in the frequency band between 0 and 1.5 kHz
- HNR25 : harmonic to noise ratio in the frequency band between 0 and 2.5 kHz
- HNR35 : harmonic to noise ratio in the frequency band between 0 and 3.5 kHz

An example of basic features extracted on a spontaneous laughter from our corpora is displayed on Figure 3.

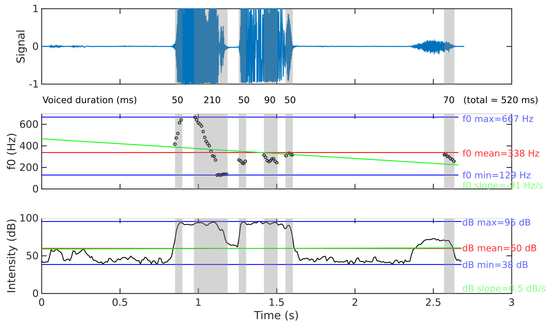

FIGURE 3: *Extraction of acoustic features on a spontaneous laughter excerpt*

4.2. Multiple Factor Analysis

To explore the global correlation between the acoustic features of F0 (mean, slope, standard deviation), intensity (mean, slope, standard deviation), total duration and voiced segment duration and the perceptual values (responses provided by the subjects) of both French and Japanese groups (abbreviated as Res_FR and Res_JP), a Multiple Factor Analysis (MFA) was carried out. Before computing the MFA, all acoustic and perceptual values were converted into z-scores setting average value as reference value for each parameter. The result showed that the distribution of the responses of French as well as for Japanese listeners were correlated with F0 features (mean and standard deviation) and the total duration of the laughter segments and of the voiced segments. However, the intensity (mean, slope, standard deviation) and F0 slope were less correlated with the perceptual responses of the two groups (Figure 4). Table 2 shows the values for F0 mean, F0 sd, mean duration and voiced segment mean duration. Significant differences were found between spontaneous and social laughter for F0 sd (t(52)=5.669, p.=0.05), for duration mean (t(52)=2.696, p.=0.05) and the voiced segment duration mean (t(52)=2.595, p.=0.05) between spontaneous and social volitional laughs. The variations of F0 values are higher, total duration and voiced segment duration is longer for spontaneous laughs than for social ones.

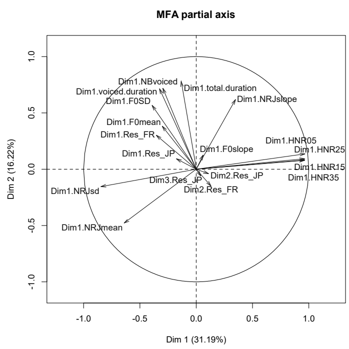

FIGURE 4: *Correlation between acoustic and perceptual values described by Multiple Factor Analysis*

TABLE 2: *Mean F0, F0 SD, total duration mean and voiced duration mean for the spontaneous and the social laughs*

	Spontaneous	**Social**	**t-test**
F0 mean (Hz)	203.59	160.80	ns
F0 SD	54.75	25.69	2.696*
Total duration mean (s)	1.81	0,65	5.669**
Voiced duration mean (s)	0.25	0.13	2.595*
*$p<.05$	**$p<.01$		

4.3. Principal Component Analysis

Previous MFA analysis showed only the global correlation between all responses and all acoustic parameters. Therefore, a Principal Component Analysis (PCA) was applied to all acoustic parameters by the two types of stimuli that were categorized by all of the listeners (French and Japanese groups). We first analysed all types of laughs for the intensity (mean, slope, standard deviation) and total duration. Ellipses indicate a normal probability (=68%) for each group of laughter. Correlations are found between the intensity standard deviation and the intensity mean vectors. The direction of the vector corresponding to the total duration on the component 2 (vertical axis) reveals that these acoustic features help differentiate spontaneous laughs from the social ones (Figure 5).

A second PCA was applied on the voiced laughs only (6 completely unvoiced laughs were removed from the set) in order to add the acoustic features related to voicing to the analysis : F0 mean, F0 slope, F0 standard deviation, voicing duration, number of voiced segments. The result (Figure 6) shows that the voiced duration and the number of voiced segments are correlated. F0 standard deviation and total duration are closely correlated. Then, F0 mean and intensity slope are correlated as well. According to the distribution of the type of laughs related to the direction of each vector on the component 1, it was found that the acoustic features concerning the voiced segment duration, the number of voiced segments, the total duration and the F0 standard deviation help differentiate spontaneous and social laughs.

Figure 7 represents the variations in duration of each laughter (normalised values). Spontaneous laughs show greater variability than social laughs.

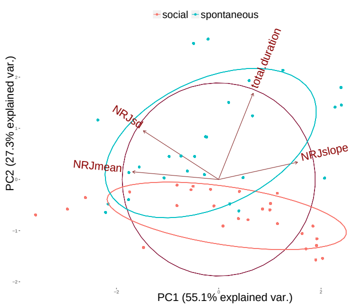

FIGURE 5: *Correlation between the acoustic values (intensity and total duration) and the 54 laughs*

5. Conclusion

The current paper investigates whether human beings can perceptually discriminate between social volitional laughter and spontaneous involuntary laughter from a corpus of spontaneous laughs recorded in a virtual immersive environment using only sound information without any context or any foreign language skill. According to the perceptual discrimination experiment with native French and Japanese subjects, participants are able to discriminate these two types of laughter indicated by more than twice the chance level of recognition rate without context. This result confirms the existence of two types of laughter on the voluntary-involuntary control dimension as mentioned in previous research [9, 17].

Fourteen acoustic features including F0, harmonic to noise ratio, intensity and duration for each type of laughter are also investigated. Multiple factor Analysis was conducted to explore the global correlation between the acoustic characteristics and the participants' perceptual behaviour. Results showed that the perceptual behaviours of both French and Japanese groups were correlated with F0 features (mean and standard deviation), the total duration and the voiced segment duration. After this global result, we further investigated the important acoustic factors associated to each type of laughter (spontaneous or social). The results showed that the total duration helps to differentiate spontaneous laughs from the social ones. Moreover, we found that the voiced duration, the number of voiced segments and the F0 standard deviation also contribute to the differentiation between spontaneous and social laughs.

For future work, we will implement an additional perceptual experiment with social laughs to explore sub-categories of social laughter (embarrassment, politeness or mirthful) among two different cultures/languages (i.e. Japanese and French).

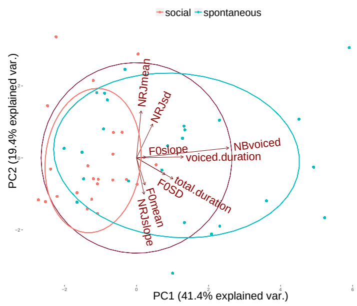

FIGURE 6: *Correlation between the acoustic values (F0, intensity, total duration, voicing duration) and the 48 laughs*

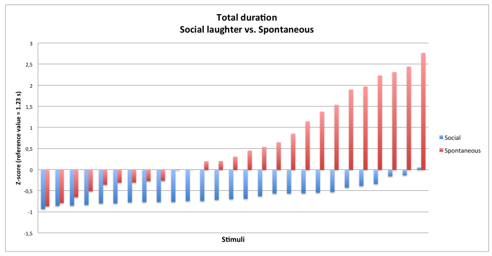

FIGURE 7: *Total duration of the two types of laughs*

6. Acknowledgements

This study has been carried out with financial support from the French State in the frame for the "Investments for the future" program IdEx Bordeaux (ANR-10-IDEX-03-02), the Bordeaux PEPS IDEX/CNRS project "Virtual Laughter", the Bordeaux International Support in collaboration with Tsukuba University

"Cultural difference of Social Laughter". We also wish to thank all the speakers and listeners from Kyoto University and Bordeaux University for their kind participation.

7. References

[1] H. Ishiguro, "Communication robots," in *International Congress of Psychology*, Yokohama, Japan, 2016.

[2] T. Shochi and A. Rilliard, "La prononciation des apprenants de fle et la multimodalite expressive," in *La prononciation du français dans le monde*. CLE International, 2016, pp. 257–263.

[3] A. Wichmann, "The attitudinal effects of prosody, and how they relate to emotion," in *International Speech Communication Association Tutorial and Research Workshop on Speech and Emotion*, 2000.

[4] N. Campbell, H. Kashioka, and R. Ohara, "No laughing matter." in *Interspeech*, Lisbon, Portugal, 2005.

[5] R. R. Provine, "Laughter," *American Scientist*, vol. 84, pp. 38–45, 1996.

[6] ——, *Laughter : A scientific investigation*. Penguin, 2001.

[7] M. Owren and J. Bachorowski, "Acoustic assessment of vocal expression of emotion," *Handbook of emotion elicitation and assessment. Oxford University*, pp. 239–266, 2007.

[8] D. Erickson, C. Menezes, and K.-i. Sakakibara, "Are you laughing, smiling or crying ?" in *Asia-Pacific Signal and Information Processing Association Annual Summit Conference*, Sapporo, Japan, 2009.

[9] S. K. Scott, N. Lavan, S. Chen, and C. McGettigan, "The social life of laughter," *Trends in cognitive sciences*, vol. 18, no. 12, pp. 618–620, 2014.

[10] T. Jacykiewicz and F. Ringeval, "Automatic recognition of laughter using verbal and non-verbal acoustic features," Ph.D. dissertation, Masters thesis, Department of Informatics, University of Fribourg, Switzerland, 2014.

[11] B. Wild, F. A. Rodden, W. Grodd, and W. Ruch, "Neural correlates of laughter and humour," *Brain*, vol. 126, no. 10, pp. 2121–2138, 2003.

[12] J.-A. Bachorowski, M. J. Smoski, and M. J. Owren, "The acoustic features of human laughter," *The Journal of the Acoustical Society of America*, vol. 110, no. 3, pp. 1581–1597, 2001.

[13] H. Tanaka and N. Campbell, "Acoustic features of four types of laughter in natural conversational speech," *International Congress of Phonetic Sciences*, 2011.

[14] A. Anikin and C. F. Lima, "Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations," *The Quarterly Journal of Experimental Psychology*, pp. 1–21, 2017.

[15] S. Kipper and D. Todt, "Variation of sound parameters affects the evaluation of human laughter," *Behaviour*, vol. 138, no. 9, pp. 1161–1178, 2001.

[16] J. Vettin and D. Todt, "Human laughter, social play, and play vocalizations of non-human primates : an evolutionary approach," *Behaviour*, vol. 142, no. 2, pp. 217–240, 2005.

[17] N. Lavan, S. K. Scott, and C. McGettigan, "Laugh like you mean it : Authenticity modulates acoustic, physiological and perceptual properties of laughter," *Journal of Nonverbal Behavior*, vol. 40, no. 2, pp. 133–149, 2016.

[18] P. Boersma and W. David, *Praat, a system for doing phonetics by computer [Computer program], Version 5.3.51*, 2013.

[19] R. Jürgens, K. Hammerschmidt, and J. Fischer, "Authentic and play-acted vocal emotion expressions reveal acoustic differences," *Frontiers in psychology*, vol. 2, 2011.

[20] S. Mathôt, D. Schreij, and J. Theeuwes, "Opensesame : An open-source, graphical experiment builder for the social sciences," *Behavior research methods*, vol. 44, no. 2, pp. 314–324, 2012.

[21] R Core Team, *R : A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. [Online]. Available : https ://www.R-project.org

[22] K. Sjolander, *The Snack Sound Toolkit [computer program]*, 2000.