

VOCAL ASPECT OF SOCIAL LAUGHTER DURING VIRTUAL INTERACTION

ASPECTOS VOCAIS DO RISO SOCIAL NA INTERAÇÃO VIRTUAL

Takaaki Shochi | [CLLE Curriculum](#) | takaaki.shochi@labri.fr
CLLE-ERSSàB UMR5263 & LaBRI CNRS UMR 5800, Bordeaux, France

Marine Guerry | [CLLE Curriculum](#) | marine.guerry@gmail.com
CLLE-ERSSàB CNRS UMR5263, Bordeaux, France

Hanako Suzuki | [Ritsumeikan University Curriculum](#) | suzukih@fc.ritsumei.ac.jp
Ritsumeikan University, Japan

Mami Kanzaki | [J-Global](#) | kanmami.224@gmail.com
Ritsumeikan Global Innovation Research Organization, Japan

Jean-Luc Rouas | [LaBRI Curriculum](#) | jean-luc.rouas@labri.fr
LaBRI, CNRS UMR 5800, Talence, France

Toyoaki Nishida | [Nishida Lab](#) | nishida@i.kyoto-u.ac.jp
Kyoto University, Japan

Yoshimasa Ohmoto | [J-Global](#) | ohmoto@i.kyoto-u.ac.jp
Kyoto University, Japan

Resumo: O presente artigo enfoca os diversos tipos de riso registrados durante interações sociais reais realizadas em um ambiente de imersão virtual. Neste experimento, investigamos se os seres humanos são capazes de discriminar risos sociais de risos espontâneos e gargalhadas audiovisuais, sem terem acesso a nenhum contexto de produção do riso. Para tal, realizamos dois experimentos perceptuais considerando apenas o estímulo auditivo, por um lado, e o estímulo audiovisual, por outro. Os sujeitos dos testes de percepção são falantes nativos franceses e japoneses. Cada sujeito ouviu e ouviu e viu 162 risadas e escolheu uma resposta entre três possibilidades: social, espontânea ou irreconhecível. Os resultados de ambos os experimentos mostram que todos os participantes são capazes de discriminar esses dois tipos de riso com um grau de confiança bastante elevado, mesmo

sem as informações contextuais: a taxa de identificação correta para o riso espontâneo é de cerca de 70%, com uma quantidade semelhante de riso social considerando tanto os estímulos visuais quanto os audiovisuais. A seguir, foram extraídas medidas acústicas para cada risada, a fim de investigar possíveis diferenças entre os dois tipos de riso. A análise multifatorial demonstra que os comportamentos perceptuais e algumas características acústicas (F0 e duração) estão correlacionados. Observamos, particularmente, uma diferença significativa entre o riso social e o riso espontâneo através dos traços da duração total e da duração do vozeamento. Por último, realizamos mais um experimento perceptual sobre a subcategorização dos risos sociais com base em três fatores sociais: estado físico do falante, envolvimento do falante e distância psicológica. Os resultados mostram que o riso social é evocado de forma semelhante no que diz respeito ao estado físico do falante e envolvimento do falante. Entretanto, no que diz respeito à distância social, há divergência de percepção entre franceses e japoneses sendo que nestes últimos predomina a percepção da distância psicológica.

Palavras-chave: Riso; Reconhecimento; Análise acústica; Percepção; Cultura.

Abstract: The current paper focuses on the various types of laughter recorded during real social interactions in a virtual immersive environment. With this experiment, we investigate whether human beings are able to discriminate social from spontaneous laughter on the basis of auditory or audiovisual laughs presented outside any context. Towards this aim, we carried out two perceptual experiments proposing audio alone and audiovisual conditions, which were taken by French and Japanese native subjects. Each subject listened to (or looked at) 162 laughs and chose one response among three possibilities: social, spontaneous or unknown. The results of both experiments show that all participants 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 in audio alone and audiovisual conditions. We then extracted acoustic characteristics for each laugh in order to investigate potential differences between the two types of laughter. A multifactorial analysis showed that perceptual behaviors and some acoustic features (F0 and duration) are correlated. Especially, we observe a significant difference between social and spontaneous laughter through the features of total duration and voiced duration. Finally, we conducted a perceptual experiment on the subcategorization of social laughs based on three social factors: speaker's

physical state, speaker's involvement and psychological distance. The results show that social laughter is characterized by similar contexts for both groups of listeners, except by Japanese subjects who regard the psychological distance between partners as distant.

Keywords: 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 the cognitive processing of various affective meanings as well as technical realization of such expressions. However, the 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 (ISHI et al., 2016; SHOCHI et al., 2016).

Human social interaction consists in an exchange of social information conveyed by voice, eye contact, gestures, facial expressions, sighs or laughter (SHOCHI et al., 2016; WICHMANN, 2000; CAMPBELL et al., 2005). Among these modalities, laughter is probably one of the most important behaviors in the development of speech and in human and animal communication (PROVINE, 1996; PROVINE, 2001; OWREN; BACHOROWSKI, 2007). Laughter is often considered as a physical reaction to external stimuli, which is often linked to positive valence (i.e. joyful reaction). In Jacykiewicz and Ringeval (2014) the authors report 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 from the hypothalamus (WILD et al., 2003), including muscular inputs and the respiratory axis.

According to the field of Gelotology¹, laughter is classified under 15 categories (BERK, 2001). The first three ones: “smirk”, “smile” and “grin” are perfectly voluntarily controllable without a laugh sound. Both “snicker” and “giggle” use facial muscles as “grin” but are accompanied by laugh sounds. “Chuckle” involves chest muscles with deeper pitch, and “chortle” involves muscles of the torso. This type of laugh usually provokes laughter in others (BERK, 2001). When a laugh is accompanied by barking or snorting, it is usually classified in the category of “laugh [which] involves facial and thoracic muscles as well as abdomen and extremities” (BERK, 2001, p. 327). All these eight types of

¹ “Gelotology” from the Greek “γέλως / *gelos*” (meaning “laughter”) and “λογία / *logia*” (study of). (BUTLER, 2005).

laughs are described as controllable and can be voluntary simulated. On the contrary, seven other laughs: “cackle”, “guffaw”, “howl”, “shriek”, “roar”, “convulse” and “die laughing” are characterized by an uncontrollable and involuntary nature. This physical description of laughter is deeply rooted in human biology. Therefore, these characteristics should apply to every human being. However, laughter also serves very important social roles to bring about positive, mutually beneficial relationships among people and communities (ERICKSON et al., 2009; SCOTT et al., 2014). When laughter plays social roles, it must be strongly influenced by the culture we belong to. For instance, R.P. Luís Fróis, a Portuguese missionary who travelled in Asia in the 16th century, mentioned cultural differences in the use of laughter between Portuguese and Japanese in his manuscript “About the difference between Europe and Japan”.

“In our country, a feigned laugh is considered hypocritical; In Japan, it is a mark of elegance and good disposition”

(Luís Fróis, 1585 p. 87)

“In our country, politeness is made of a serene and grave attitude; The Japanese infallibly do so with all sorts of little simulated laughs”

(Luís Fróis, 1585 p. 89)

These descriptions suggest two important pieces of information about laughter. Firstly, laughs are not always linked to amusement: people frequently simulate laughs, which play an important social role in face-to-face communication. Secondly, there is a clear perceptual gap between laughter of two cultures. Such cultural differences influence how people laugh depending on various social contexts.

Indeed, a few studies have already tried to reveal the cultural aspects of laughter, especially focusing on the distinction between two types of laughs: social laughter and an amusing (or mirthful) one (PROVINE, 2001; VETTIN; TODT, 2005; ERICKSON et al., 2009; TANAKA; CAMPBELL, 2011).

In Scott et al. (2014), the authors suggest the existence of two different types of laughter: spontaneous and volitional (or social), distinguished 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, recruited in order to set up a positive relationship or to tone down a conflictive tension.

Concerning the acoustic realization of these various types of laughter, Trouvain (2003) summarizes an analysis of various types of laughter in three levels: “segmental”, “syllable” and “phrasal”. With regard to acoustic analyses at the phrasal level, Tanaka; Campbell (2011, 2014) made an acoustic distinction between “mirthful” involuntary spontaneous laughs and “politeness” laughs which is closely linked to social relationships; spontaneous mirthful laughs tend to include chuckles and ingressive sounds which were rare in polite laughs. Moreover, mirthful involuntary laughs tend to be longer than polite social laughs. Other recent works (ANIKIN; LIMA, 2017; BACHOROWSKI et al., 2001; KIPPER; TODT, 2001; VETTIN; TODT, 2005) support the idea that spontaneous laughter has a higher and a more variable fundamental frequency (hereby F0), as well as a higher variability in acoustic parameters in general. In addition, spontaneous laughter is also characterized by a longer duration with shorter bursts, ingressive and chuckle sounds (TANAKA; CAMPBELL, 2011; LAVAN et al., 2016). However, there is no significant difference for both types of laughter regarding breathiness and mouth aperture.

According to our hypothesis, (1) human beings are able to discriminate between social (voluntary) laughs and mirthful spontaneous (involuntary) ones using only audio cues without visual or contextual information; (2) perceptually determined mirthful involuntary laughter may have common acoustic cues among different cultures; and on the contrary, (3) volitional social laughter may be perceived differently from one culture to another due to cultural conventions.

Following these hypotheses, the current research investigates (1) whether French and Japanese subjects can discriminate between social volitional laughter and spontaneous involuntary laughter using a) only auditory laughs or b) audiovisual laughs, extracted from an immersive virtual interaction environment, and without any context. Independently we aim at investigating (2) the acoustic characteristics for each type of laughter in French and Japanese. Moreover, (3) we try to describe a socio-linguistic context for a majority of social volitional laughter.

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 communication between people in virtual environments and was made using Unity². The game was designed by two scenario designers. They discussed how to control the virtual characters in the recording scenario for inducing spontaneous affective reactions in players. The motions of the virtual characters were made from interaction motions of humans captured by a Kinect sensor so to provide natural motions. The scenario designers carefully arranged the captured motions and pre-recorded motions, which were distributed by the asset store of Unity3D. Each player was alone in his own individual immersive virtual environment (completely surrounded by displays or in an immersive dome). Inside the immersive environment, participants could look around at the virtual game world with a low cognitive load, as in the real world. In addition, each player could communicate with the others using Skype with cameras and microphones, so they could recognize their facial expressions and hear each other's voices. The Skype windows were displayed on the left side of the player. They were not obstructive when the player interacted with the virtual character. The players 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 (Figure 1). A total of 12 spontaneous affective speech data files of 9 Japanese (2F/7M) and 3 French (1F/2M) were recorded. All participants were university students recruited at Kyoto, Japan. All Japanese native subjects grew up in Japan (Kansai region), and the 3 native French speakers were foreign students at Kyoto University, but were not able to communicate in Japanese at the moment of the experiment.

A total of 254 sequences containing only laughter were manually segmented using PRAAT (BOERSMA; DAVID, 2013). A first pilot test was conducted in order to investigate which acoustic features distinguish mirthful spontaneous vocalizations of laughs from volitional forms which are considered as social laughs (LAVAN et al., 2016; ANIKIN; LIMA, 2017; JÜRGENS et al., 2011). 7 experimenters (3 Japanese males and 4 French (3F/1M) were instructed to listen and 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. These 27 laughs consist in 6 laughs uttered by French speakers and 21 laughs from Japanese speakers. Another pilot test designed to select the 27 volitional social laughs produced by 2 French and 25 Japanese speakers was done under the same criterion as for the spontaneous one by the same experimenters as for the first pilot test.

² Unity is a cross-platform game engine developed by Unity Technologies.

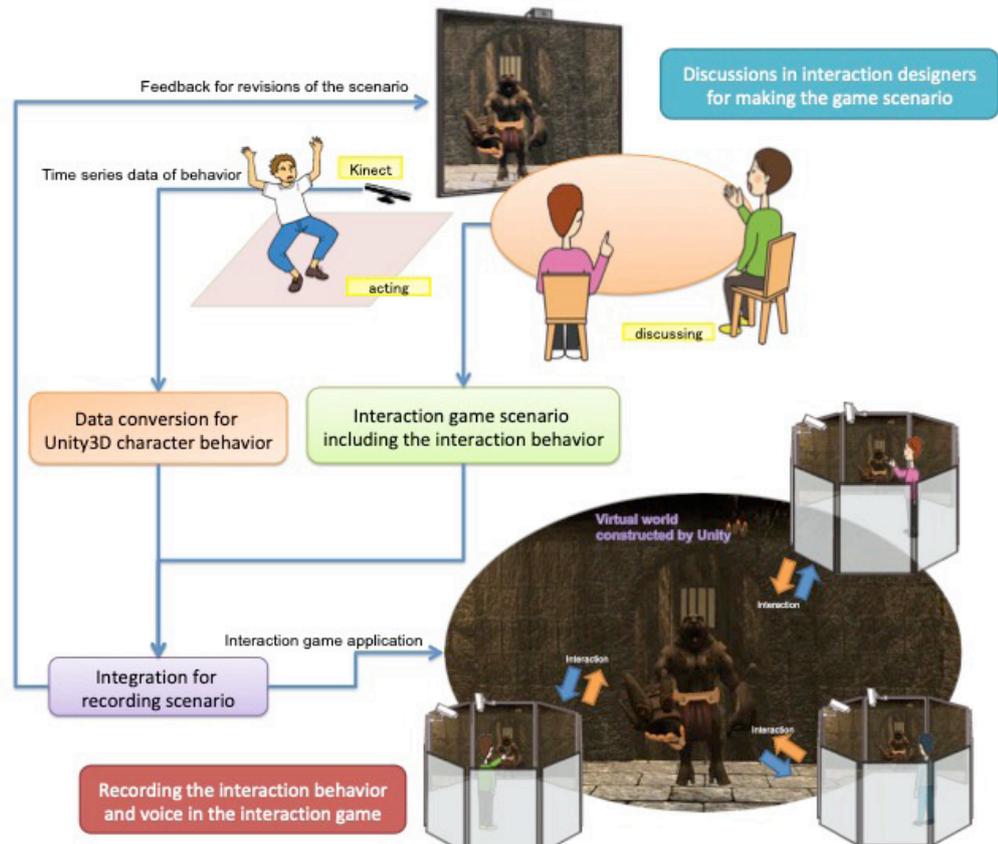


Figure 1. The immersive environment setting used in the current experiment.

3 Perceptual experiment in the audio modality

3.1 Paradigm

82 native French listeners (48F/34M, Mean age = 22.39 years) and 97 native Japanese listeners (67F/30M, Mean age = 20.42 years) were recruited in both countries. The stimuli were displayed 3 times each in audio alone condition and presented in a randomized order (54 laughs (27 spontaneous / 27 social) x 3 (repetitions) = 162 stimuli).

Before the test, subjects were informed about the definition of each type of laughter and the procedure of the experiment (Figure 2). The test was conducted individually using a GUI-based interface developed under the “OpenSesame” software (MATHÔT et al., 2012). 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). Then, they had to select one answer among the three possible ones: “spontaneous”, “social”, “I don’t know” (Figure 3). 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...)

Instructions

Nous vous remercions infiniment pour votre participation à notre expérience sur les rires.

Il vous est demandé d'écouter et d'identifier une série de rires produits par des locuteurs hommes et femmes.

L'expérience se déroule de la manière suivante:

Pour chaque rire:

1) Vous allez entendre la séquence une première fois

2) Vous devez choisir UNE réponse parmi les trois choix possibles:

- "Spontané": il vous semble que la personne rit de manière spontanée à un événement extérieur (Ex: une vidéo amusante)
- "Social": il vous semble que la personne rit pour maintenir le dialogue avec son (ses) interlocuteur(s). (Ex: rire embarrassé, rire poli, cynisme...)
- "Je ne sais pas"

3) Dans le cas où:

- vous avez sélectionné "Spontané" ou "Social". Veuillez indiquer à quel point vous êtes certain(e) que ce rire correspond à votre choix en cliquant sur une échelle de 1 (pas certain(e)) à 7 (tout à fait certain(e))
- vous avez sélectionné "Je ne sais pas", le rire suivant est lancé

ATTENTION:

Vous avez la possibilité de réécouter le son 1 SEULE FOIS SUPPLÉMENTAIRE. Une fois que vous avez appuyé sur "Écoute" pour écouter une seconde fois le stimulus, vous devrez faire votre choix pour passer au rire suivant

Veuillez suivre votre première intuition lors de votre choix.

Maintenant, vous allez commencer une phase d'entraînement afin de vous habituer à l'interface.
Vous passerez ensuite à la phase expérimentale.

Appuyez sur "Ok" pour commencer.
(le premier rire va se lancer directement)

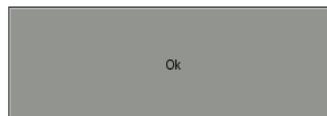


Figure 2. Part of the instructions given to the French listeners.
The interface was translated in Japanese for the Japanese listeners.

Quel est ce rire?

spontané

social

jenesaispas

Figure 3. Interface displaying the multiple choice.
From top to bottom: "spontaneous", "social", "I don't know"

3.2 Results

First of all, a χ^2 test was computed on all the responses for both groups of listeners (French and Japanese) to check whether their distributions (Social, Spontaneous or Unknown) are independent. According to the result, a significant difference for the distribution of answers was observed ($\chi^2 = 9396.5$, $df: 2$, $p < .001$).

According to Table 1 and Table 2 (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 correctly 69.24% of spontaneous laughs and 69.41% of social laughs; Japanese listeners recognized 72.25% of spontaneous laughs and 69.02% of social laughs. This result confirms that the listeners of both groups are able to recognize two types of laughter without visual indices or context.

Table 1. Results for the perceptual test for 82 French listeners.

Raw results are presented with their frequency for each row.

	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%)

Table 2. Results for the perceptual test for 97 Japanese listeners.

Raw results are presented with their frequency for each row.

	Spontaneous	Social	Unknown
Spontaneous	5677 (72.25%)	1908 (24.28%)	272 (3.46%)
Social	1219 (15.51%)	5423 (69.02%)	1215 (15.46%)
Total result	6896 (43.88%)	7331 (46.65%)	1487 (9.46%)

3.3 Distribution of the perceptual values

In order to observe the perceptual distance of all responses based on the listeners' classification (spontaneous, social, I don't know) for the 54 stimuli, we computed a Correspondence Analysis (CA) for both Japanese and French groups using FactoMineR package (R project) under R software. According to the CA, the perceptual behavior for 26 stimuli in the French group and 20 stimuli in the Japanese one, listeners showed an important contribution (i.e. above the expected average contribution) on the first two dimensions which explain 99.8% of variances for French group and 100% of variances for Japanese group. The blue dots on the figures represent the distribution of the perceptual behavior and the three red triangles represent the concept subjects have of the types of laughter. Therefore, the closer the blue dots are to the red triangle, the more it is perceived as this type of laughter.

Similarly to the global observation, this statistical analysis shows the perceptual patterns in French and Japanese groups. In Figure 4 and Figure 5, the perceptual cluster of "sponta-

neous” laughter is located in mid left side, while the cluster for “social” laughter is located on the right bottom corner. It indicates that French and Japanese groups are able to discriminate clearly the two types of laughter showing some confusion between “social” and “unknown”.

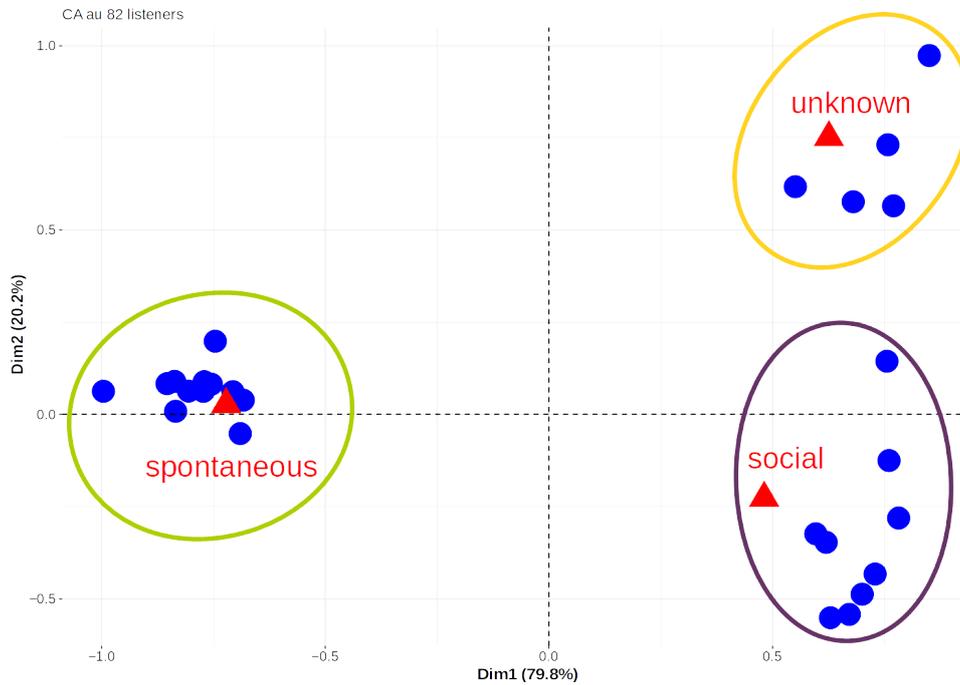


Figure 4. Distribution of the perceptual behavior of the French listeners for 26 stimuli - audio alone condition

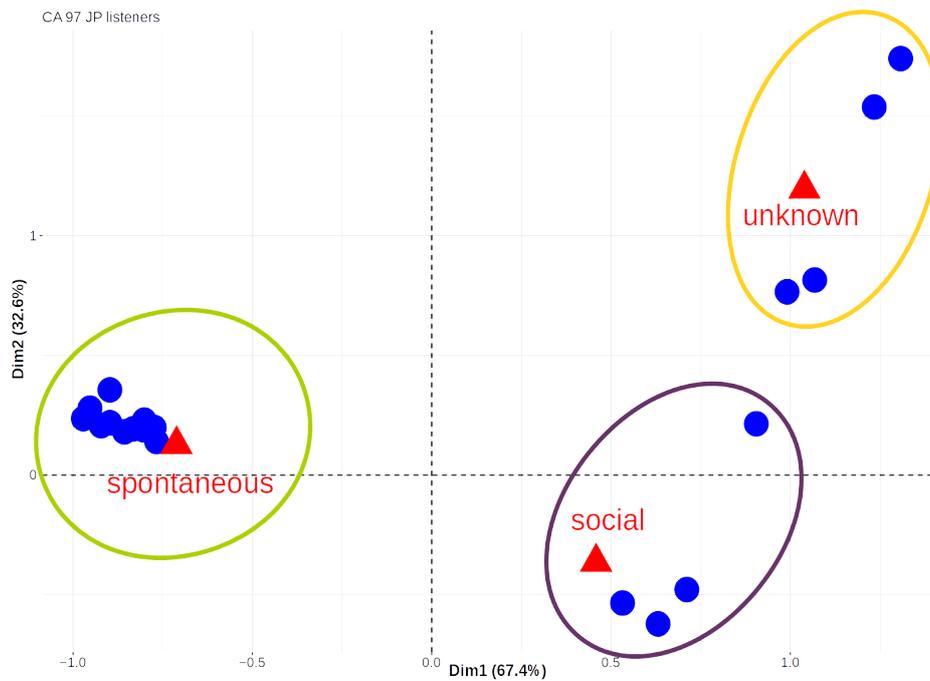


Figure 5. Distribution of the perceptual behavior of the Japanese listeners for 20 stimuli - audio alone condition

4 Perceptual experiment in the audiovisual modality

4.1 Paradigm

98 French native listeners (65F/33M, Mean age = 20.64 years) and 87 Japanese native listeners (65F/22M, Mean age = 19.74 years) were recruited in both countries. The stimuli were displayed 3 times each during the audiovisual condition in a randomized order (54 laughs (27 spontaneous / 27 social) x 3 (repetitions) = 162 stimuli). The stimuli are the same as those used for the audio alone condition. The instructions were also identical to the audio alone condition.

4.2 Results

A χ^2 test was computed on all the responses for both groups of listeners (French and Japanese) to evaluate whether the distributions of listeners' responses (Social, Spontaneous or Unknown) are independent. According to the result, a significant difference in the distribution of answers was observed between the two groups ($\chi^2 = 4225.1$, $df = 2$, $p < 0.001$).

According to Tables 3 and Table 4, the two types of laughter are fairly recognized: French subjects correctly identified 61.64% of spontaneous laughs and 70.55% of social laughs; Japanese listeners recognized 69.39% of spontaneous laughs and 73.43% of social laughs. These results confirm that the listeners of both groups were able to recognize the types of laughter with the audiovisual as well as with the audio stimuli. Unlike the results on audio stimuli, the "social" laughs seem to be better discriminated than the "spontaneous" one for both groups. This could suggest that the subjects perceived more "social" information from visual cues.

Table 3. Results for the perceptual test for 98 French listeners.

Raw results are presented with their frequency for each row

	Spontaneous	Social	Unknown
Spontaneous	4893 (61.64%)	1715 (34.20%)	330 (4.15%)
Social	1672 (21.06%)	5601 (70.55%)	665 (8.37%)
Total result	6565 (41.35%)	8316 (52.38%)	995 (6.26%)

Table 4. Results for the perceptual test for 87 Japanese listeners.

Raw results are presented with their frequency for each row

	Spontaneous	Social	Unknown
Spontaneous	2245 (69.39%)	905 (27.94%)	86 (2.65%)
Social	663 (20.52%)	2372 (73.43%)	195 (6.03%)
Total result	2908 (44.98%)	3276 (50.67%)	281 (4.34%)

4.3 Distribution of the perceptual value

As for the responses obtained on audio stimuli, we performed the same statistical analysis on answers from the audiovisual condition. According to the CA, the perceptual behavior for 21 stimuli in the French group and 20 stimuli in the Japanese one among 54 stimuli showed an important contribution (i.e. above the expected average contribution) on the first two dimensions (100% of explained variance for both groups).

This CA analysis shows that both French and Japanese listeners have a similar perceptual behavior. In Figures 6 and 7, the perceptual groups of “spontaneous” and “social” laughs are opposed to each other on the 1st dimension. It indicates that both listeners groups are able to discriminate the two types of laughter. However, they show confusions between “social” and “unknown” for the same stimuli.

These results from audiovisual stimuli follow mostly the same perceptual pattern as the audio only condition. Therefore, it confirms that the visual information already provides an important amount of cognitive information in order to be able to discriminate two types of laughter.

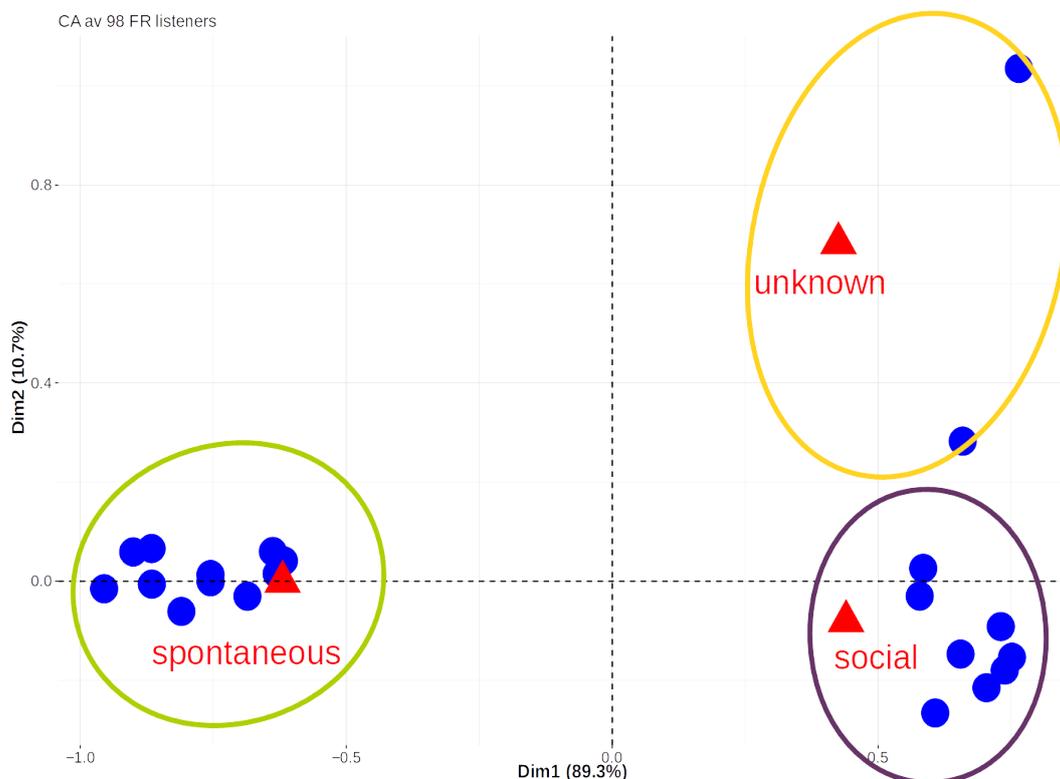


Figure 6. Distribution of the perceptual behavior of the French listeners for 21 stimuli - audio visual condition

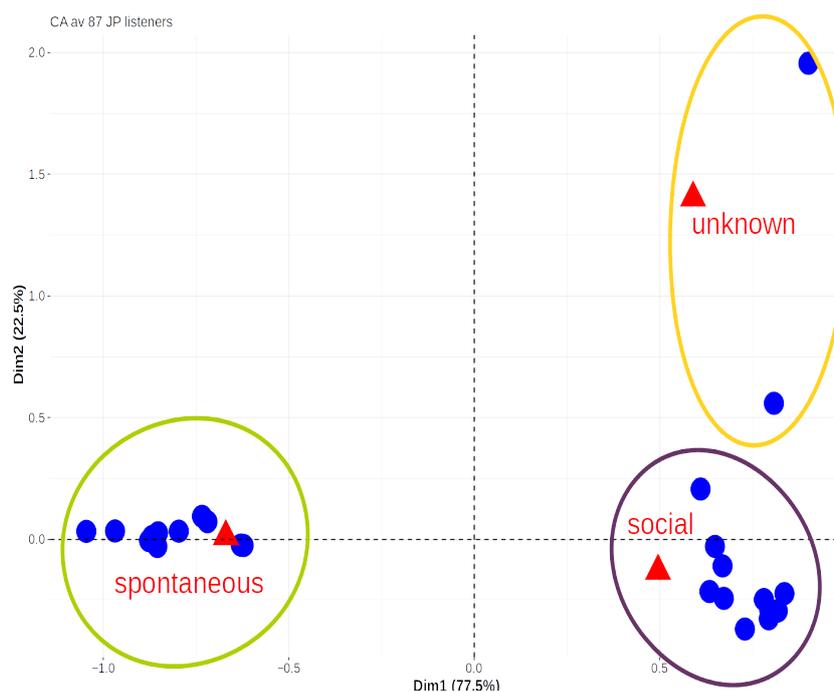


Figure 7. Distribution of the perceptual behavior of the Japanese listeners for 24 stimuli - audio visual condition

5 Acoustic analysis

Since our initial categorization of two types of laughter (mirthful spontaneous laughter and social one) is validated by both perceptual experiments (audio alone and audiovisual stimuli), we measure in the next section several acoustic features that were reported to predict affective ratings and categorization for laughter as well as for more general affective voice analysis in previous research (LAVAN et al., 2016).

5.1 Features extraction

For the acoustic analysis, the fundamental frequency (F0) and global intensity are computed every 10 ms. They are extracted using a customized version of the Snack toolkit (SJOLANDER, 2000). Most analyses are carried out on the voiced parts of the laughter as detected by the F0 extraction algorithm, thus ignoring voiceless segments.

We extracted a set of 13 features for each laugh in four main categories: F0 values for assessing the variability of the F0 (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. The details of the features are:

- The mean value of F0 (F0 mean) extracted on voiced parts of the laughs (Hz)
- The standard deviation of F0 values (F0 sd) on a laughter excerpt (voiced parts)
- The approximated F0 slope (voiced parts only)
- The F0 minimum (F0 minimum)
- The F0 maximum (F0 maximum)
- The F0 range (F0 range)
- The mean of intensity (NRJ mean) values (dB)
- The standard deviation of intensity values (NRJ sd) during a laughter
- The intensity minimum (NRJ minimum)
- The intensity maximum (NRJ maximum)
- The intensity range (NRJ range)
- The total duration of a manually segmented laughter
- The duration of all the voiced parts of a laughter
- The percentage of voiced segments during a laughter
- The number of voiced segments
- The harmonic to noise ratio (HNR) in the frequency band between 0 and 0.5 kHz
- The harmonic to noise ratio (HNR) in the frequency band between 0 and 1.5 kHz
- The harmonic to noise ratio (HNR) in the frequency band between 0 and 2.5 kHz
- The harmonic to noise ratio (HNR) in the frequency band between 0 and 3.5 kHz

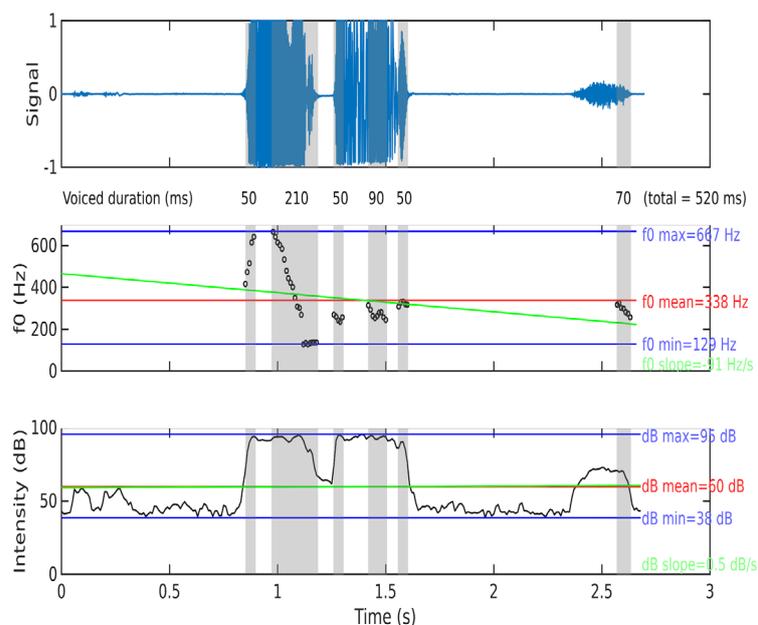


Figure 8. Extraction of acoustic features on a spontaneous laughter excerpt

5.2 Multiple Factor Analysis

To explore the global correlation between the acoustic features of F0 (mean, slope, standard deviation (sd)), 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 in audio alone condition, a Multiple Factor Analysis (MFA) was carried out. Before computing the MFA, all acoustic and perceptual values were converted into z-scores by setting the average value as reference value for each parameter. Figure 9 illustrates arrows which represent vectors of acoustic and perceptual variables. According to the MFA analysis, arrows for acoustic and perceptual variables having similar directions are considered as positively correlated variables, and are grouped together, whereas negative ones are positioned on opposite sides. This analysis shows that the distribution of the responses for French as well as for Japanese listeners are correlated with F0 features (mean and standard deviation) and the total duration of the laughter segments and of the voiced segments. However, the relative intensity (mean, slope, standard deviation) and F0 slope are less correlated with the perceptual responses of the two groups (Figure 9). Significant differences are 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 voiced segment duration mean ($t(52)=2.595$, $p=0.05$) between spontaneous and social volitional laughs (Table 5). The variations of F0 values are higher, total duration and voiced segment duration are longer for spontaneous laughs than for social ones.

Table 5. 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**

* $p<.05$, ** $p<.01$

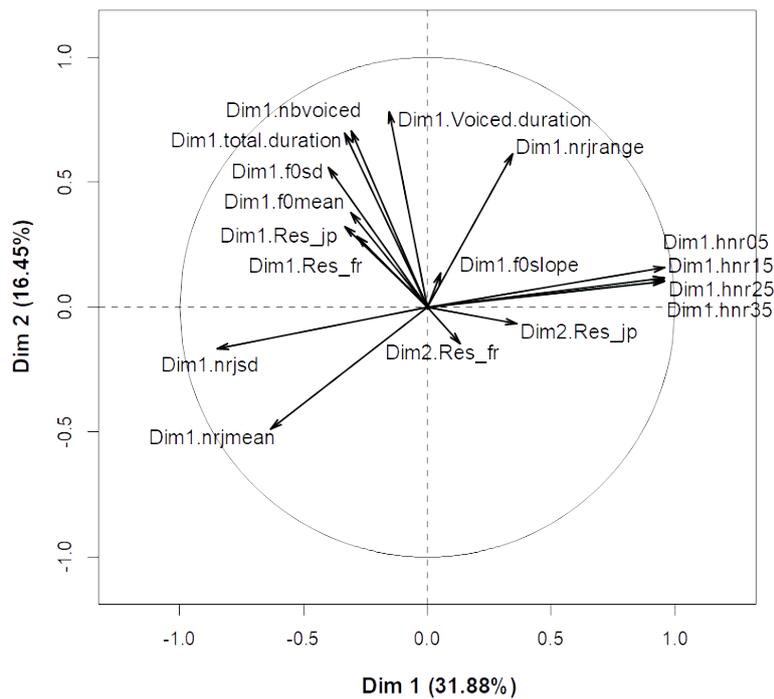


Figure 9. Correlation between acoustic and perceptual values of audio stimuli described by Multiple Factor Analysis. Vectors showing the closed direction represent positive correlation whereas negative ones are positioned on opposite sides.

5.3 Principal Component Analysis

The 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 for the two types of stimuli that were categorized by all listeners (French and Japanese groups). We first analyzed all types of laughs for the intensity (mean, slope, standard deviation) and total duration (Figure 10). According to this analysis, 1st dimension (PC1) and 2nd dimension (PC2) explain 82.4% of variances. The result shows that correlations are found between the intensity standard deviation and the intensity mean vectors.

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 sd, voicing duration, number of voiced segments (Figure 11). With regard to this graph, the 1st (PC1) and 2nd dimensions (PC2) explain 60.8% of variance. The result shows that the voiced duration and the number of voiced segments are correlated. F0 sd 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 sd help differentiate spontaneous and social laughs.

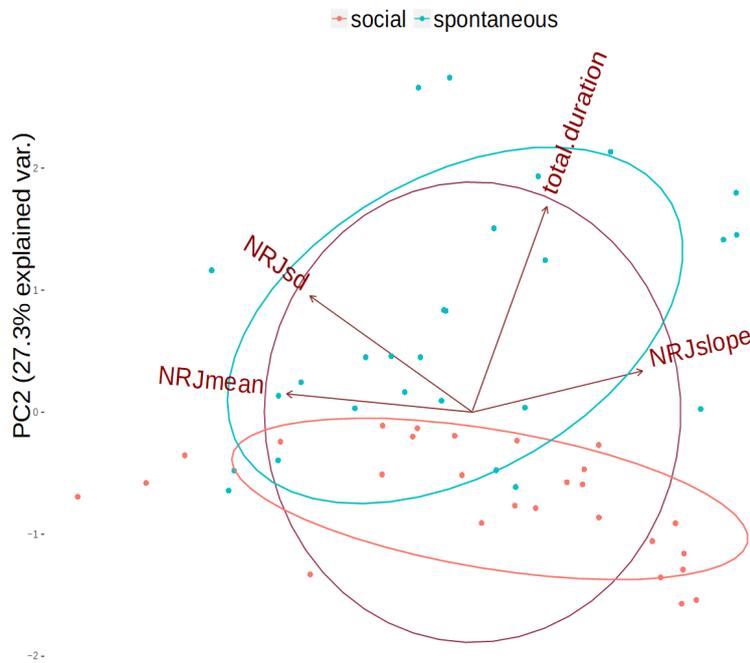


Figure 10. Correlation between the acoustic values (intensity and total duration) and the 54 laughs

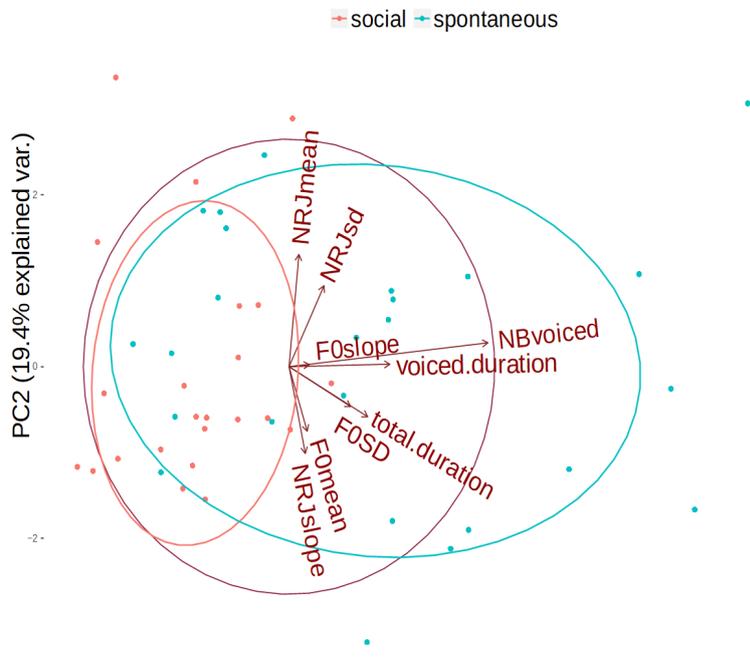


Figure 11. Correlation between the acoustic values (F0, intensity, total duration, voicing duration) and the 48 laughs

6 Subcategories of social laughter

6.1 Method

After the previous results which confirmed the initial hypothesis: human being are able to discriminate between social (voluntary) laughs and spontaneous (involuntary) ones using only audio information, we focused on the perceptual subcategorization of social laughs relying on the various psychological dimensions (LIBERMAN et al., 2007; TIAN et al., 2016; SOKOLOV; BOUCSEIN, 2000), and which may be different among languages/cultures.

Following some representative works in the field of social psychology and linguistics, three social factors: 1) speaker's physical state, 2) speaker's involvement, 3) psychological distance between speaker and partner are set up in order to create the social context in which these various social laughs occur. Each factor is evaluated on using 7 points scales on 4 dimensions as presented below:

- Speaker's physical state: calm (1) – excited (7) (dimension: arousal)
- Speaker's involvement: cooperative (1) – non-cooperative (7) (dimension: cooperation)
- Psychological distance: 2 sub-factors:
 - Close (1) – distant (7) (dimension: proximity)
 - Superior (1) – inferior (7) (dimension: hierarchy)

6.2 Experimental procedure

6.2.1 Subjects

A number of 50 French subjects (mean age = 25.9, SD= 5) and 12 Japanese subjects (mean age = 37.0, SD = 6.06) participated in the judgment task.

6.2.2 Corpus instruction

Among 27 social laughs, which were used in the previous experiment, and 14 social laughs which were best recognized by native listeners were selected as stimuli for the current perceptual experiment. An experimental interface was developed under Google form in audio alone condition. Each participant was presented with all 14 laughs in randomized order for each of 4 dimensions separately. For each factor, they were required to listen to the stimulus and rate the sound on each scale.

6.3 Results

First of all, a Kendall's coefficient of concordance was computed on the responses in order to assess the listeners' concordance of their rating each factor. The results (Table 6, Table 7) show that no general tendency was identified for both groups of listeners on the evaluation of the 14 stimuli. It indicates that the concordance of the judgments for all 14 stimuli was low, and it seems that a variety of contexts, which subjects evoked for all 14 stimuli was important for both groups.

Table 6. Results of the Kendall's coefficient of concordance for 14 social laughs rated by the 50 French listeners on the 4 dimensions

Arousal: $W=0.261114$, $\chi^2 = 169.724$ (df :13, $p.<.01$) ***
Proximity: $W=0.230408$, $\chi^2 = 149.765$ (df :13, $p.<.01$) ***
Cooperation: $W=0.266099$, $\chi^2 = 172.964$ (df :13, $p.<.01$) ***
Hierarchy: $W=0.0832044$, $\chi^2 = 54.0828$ (df :13, $p.<.01$) ***

Table 7. Results of the Kendall's coefficient of concordance for 14 social laughs rated by the 12 Japanese listeners on the 4 dimensions

Arousal: $W=0.407192$, $\chi^2 = 63.5220$ (df :13, $p.<.01$) ***
Proximity: $W=0.348421$, $\chi^2 = 54.3536$ (df :13, $p.<.01$) ***
Cooperation: $W=0.330499$, $\chi^2 = 51.5579$ (df :13, $p.<.01$) ***
Hierarchy: $W=0.157175$, $\chi^2 = 24.5193$ (df :13, $p.<.05$) *

6.4 Descriptive analysis

In this section, all responses for each dimension are described and we look for a context which corresponds to the major social laughs for both groups.

6.4.1 Arousal

Concerning the results for the French subjects for the dimension of arousal, answers ranging from 1 to 3 are considered as "calm". Answer 4 is neither calm nor excited. Answers 5, 6, 7 are interpreted as "excited". The results show that 11 laughs over 14 were judged as "calm" (the sum of the answers 1, 2, 3 constitutes more than 50% of the overall given answers). The majority of the answers are distributed among answers 2 and 3. However, 3 laughs (stimulus 3, stimulus 10 and stimulus 11) were judged as "excited". These results show that French listeners have a tendency to consider the selected social laughs as being "calm".

Japanese subjects share the same tendency for 10 laughs over 14. The four other laughs were considered as “excited” (stimulus 3, stimulus 4, stimulus 8 and stimulus 11). Among those four laughs, French and Japanese subjects share the same perceptual behavior for stimulus 3 and 11 but stimulus 4 and 8 were perceived differently.

Those results show that both groups of subjects have the tendency to globally judge social laughs as “calm”.

6.4.2 Proximity

In the dimension of the social proximity, answers 1, 2, 3 represent socially “close” between the speaker and their partner. The answer 4 is neither “close” nor “distant”. Answers 5, 6, 7 represents socially “distant” between the speaker and their partner. The results for the French listeners showed that seven laughs over 14 were judged as socially “close” between speakers. A number of 3 laughs were neither “close” nor “distant” and five laughs (stimulus 9, 12, 13 and 14) were considered socially “distant”. These results suggest that French listeners have a tendency to perceive the social relationship between speakers as “close”.

However, a majority of Japanese listeners judged seven laughs socially “distant” between speakers. Two laughs were considered neither “close” nor “distant” and five laughs were considered as socially “close”. It is important to note that even if Japanese and French subjects perceive differently the dimension of proximity, they nevertheless share the same perceptual behavior for the four stimuli considered as “distant”.

6.4.3 Cooperation

For the results about cooperation, answers 1, 2, 3 are considered as “cooperative”. Answer 4 is neither “cooperative” nor “not cooperative”. Then answers 5, 6, 7 are considered as “not cooperative”. The results for the French listeners show that seven laughs over 14 are interpreted as “cooperative”. A number of three stimuli were judged as neither “cooperative” or “not cooperative” and four stimuli were considered as “not cooperative”. This result shows the tendency of the French subjects to consider the social laughs as being “cooperative”.

Similar to the French subjects, Japanese subjects judged the speaker’s attitude which appeared in the majority of social laughs (6 over 14 laughs) as being “cooperative”, three of them were interpreted as neither “cooperative” or “not cooperative”, and five laughs as “not cooperative”. Moreover, both of the groups share the same perceptual behavior by interpreting the same set of stimuli as being “not cooperative”.

6.4.4 Hierarchy

On the hierarchy dimension, answers 1, 2, 3 are considered as socially “superior” to the conversation partner. Answer 4 is interpreted as neither “superior” nor “inferior”. Answers 5, 6, 7 are interpreted as socially “inferior” to the partner. For the French group, the majority of the stimuli were considered as neither “inferior” nor “superior” with eight laughs over 14. Then, the distribution of the answers for the 6 other stimuli is equally distributed between “superior” and “inferior”. This result shows that French subjects recognized the social hierarchy between the speaker and the interlocutor as neither “superior” nor “inferior”.

As for the French group, Japanese subjects considered the social laughs as neither “superior” nor “inferior” for seven laughs over 14. More stimuli were considered as “superior” for the Japanese subjects with a number of five laughs and only two stimuli were interpreted as “inferior”. It indicates that both Japanese and French subjects share the same tendency: they globally judged the social relation between speakers as neither “superior” nor “inferior”.

7 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 auditory laughs or audiovisual laughs without any context or any foreign language skill. According to the perceptual discrimination experiment with only auditory laughs, two groups of subjects (i.e. native French and Japanese subjects) who belong to quite different cultures/languages are able to discriminate these two types of laughter indicated by more than twice the chance level of recognition rate, without visual information nor any context. In the second perceptual experiment with audiovisual laughs, both groups discriminated well showing a high recognition rate as much as in the previous audio alone condition. These results confirm the existence of two types of laughter on the voluntary-involuntary control dimension as mentioned in previous research (SCOTT et al., 2014; LAVAN et al., 2016).

Thirteen acoustic features including F0, harmonic to noise ratio, intensity and duration for each type of laughter are also investigated. It is important to mention that the acoustic variation of laughs due to the speaker’s native language (i.e. French and Japanese)

was not verified for this current experiment. Indeed, it was difficult to compare such a cultural factor because of the lack of French subjects' samples and of the importance of the intrinsic variation among all laughs. Therefore, we focused on the global analysis excluding speech variation due to the cultural factor. First of all, Multiple factor Analysis was conducted to explore the global correlation between the acoustic characteristics and the participants' perceptual behavior. Results show that the perceptual behaviors of both French and Japanese groups are 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 show 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.

We also investigated subcategories of social laughs on a subset of excerpts validated in the previous perceptual experiment. More concretely, we implemented a third perceptual experiment with 14 highly rated social laughs for Japanese and French groups. According to the psycho-dimensional analysis, we are able to describe a context for French group: the speaker knows the conversation partner. Both speaker and his partner are students. The speaker is willing to go on the conversation (and social interaction). In this context, speaker laughed in calm. Japanese subjects evoked a similar context except for the nature of psychological distance: the speaker does not know well the conversation partner. Both speaker and his partner are students. Speaker is willing to go on the conversation (and social interaction). In this context, speaker laughed in calm.

Acknowledgments

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, Ritsumeikan University and Bordeaux Montaigne University for their kind participation for their contribution.

References

- ANIKIN, A.; LIMA, C.F. Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *The Quarterly Journal of Experimental Psychology*, p. 1-21, 2017.
- BACHOROWSKI, J.A.; SMOSKI, M. J.; OWREN, M. J. The acoustic features of human laughter. *The Journal of the Acoustical Society of America*, v. 110, n. 3, p. 1581-1597, 2001.
- BERK, R. A. The active ingredients in humor: Psychophysiological benefits and risks for older adults. *Educational Gerontology*, v. 27, no 3-4, p. 323-339, 2001.
- BOERSMA, P.; DAVID, W. *Praat, a system for doing phonetics by computer* [Computer program], Version 5.3.51, 2013.
- BUTLER, B. *Laughter The Best Medicine?* Oregon Institute of Marine University of Oregon Biology, ed. Spring, v. 11, n. 1, p. 1, 2005.
- CAMPBELL, N.; KASHIOKA, H.; OHARA, R. No laughing matter. In: Interspeech 2005, Lisbon, Portugal, 2005, p. 465-468.
- ERICKSON, D.; MENEZES, C.; SAKAKIBARA, K. Are you laughing, smiling or crying? In: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, Sapporo, Japan, 2009. p. 529-537.
- FRÓIS, L., S.J. *Traité sur les contradictions et différences de mœurs écrit par le R. P. Luís Fróis au Japon, l'an 1585, traduit du portugais, Paris, ed. Chandeigne, 1998.*
- ISHI, C.; HATANO, H.; ISHIGURO, H. Audiovisual analysis of relations between laughter types and laughter motions. In : Proc. of the 8th international conference on Speech Prosody (Speech Prosody 2016). 2016. p. 806-810.
- JACYKIEWICZ, T.; RINGEVAL, F. *Automatic recognition of laughter using verbal and non-verbal acoustic features*. 2014. Ph.D. dissertation - Department of Informatics, University of Fribourg, Switzerland, 2014.
- JÜRGENS, R.; HAMMERSCHMIDT, K.; FISCHER, J. Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in psychology*, v. 2, p. 180, 2011.
- KIPPER, S.; TODT, D. Variation of sound parameters affects the evaluation of human laughter. *Behavior*, v. 138, n. 9, p. 1161-1178, 2001.
- LAVAN, N.; SCOTT, S.K.; MCGETTIGAN, C. Laugh like you mean it: Authenticity modulates acoustic, physiological and perceptual properties of laughter. *Journal of Nonverbal Behavior*, v. 40, n. 2, p. 133-149, 2016.
- LIBERMAN, N.; TROPE, Y.; STEPHAN, E. Psychological distance. In: *Social Psychology Handbook of basic principles*. Guilford Press, v. 2, 2007. p. 353-381.
- MATHÔT, S.; SCHREIJ, D.; THEEUWES, J. Opensesame: An opensource, graphical experiment builder for the social sciences. *Behavior research methods*, v. 44, n. 2, p. 314-324, 2012.

- OWREN, M.; BACHOROWSKI, J. Acoustic assessment of vocal expression of emotion. In: *Handbook of emotion elicitation and assessment*, Oxford University, 2007. p. 239-266.
- R Core Team. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing. Vienna, Austria, 2017. Available at: <https://www.R-project.org>.
- PROVINE, R. R. Laughter. *American Scientist*, v. 84, p. 38-45. 1996.
- PROVINE, R. R. *Laughter: a scientific investigation*. New York: Penguin, 2001.
- SCOTT, S. K. et al. The social life of laughter. *Trends in cognitive sciences*, v. 18, n. 12, p. 618-620, 2014.
- SHOCHI, T.; RILLIARD, A. La prononciation des apprenants de fle et la multimodalité expressive. *La prononciation du français dans le monde*, ed. CLE International, 2016. p. 257-263.
- SJOLANDER, K. *The Snack Sound Toolkit* [Computer program], 2000.
- TANAKA, H.; CAMPBELL, N. Acoustic features of four types of laughter in natural conversational speech. Proc. 17th International Congress of Phonetic Sciences (ICPhS), Hong Kong, 2011. p. 1858-1961.
- TROUVAIN, J. Segmenting phonetic units in laughter. Proc. 15th International Congress of Phonetic Sciences, Barcelona, Spain, 2003. p. 2793-2796.
- SOKOLOV, E.N.; BOUCSEIN, W. A. Psychophysiological model of emotion space. *Integrative Physiological and Behavioral Science*, v.35, n. 2, p. 81-119, 2000.
- TIAN, Y.; MAZZOCCONI, C.; GINZBURG, J. When do we laugh? In: Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 2016. p. 360-369.
- VETTIN, J.; TODT, D. Human laughter, social play, and play vocalizations of non-human primates: an evolutionary approach. *Behavior*, v. 142, n. 2, p. 217-240, 2005.
- WICHMANN, A. 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.
- WILD, B. et al. Neural correlates of laughter and humour. *Brain*, v. 126, n. 10, p. 2121-2138, 2003.



Submitted on: 14/03/2018

Accepted on: 03/06/2018