Third Interdisciplinary Workshop
on
Laughter and other Non-Verbal Vocalisations in Speech

October 26th-27th 2012
Trinity College Dublin

TRINITY COLLEGE DUBLIN
COLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH
THE UNIVERSITY OF DUBLIN

UNIVERSITY OF TWENTE.

FastNet
Focus on Actions in Social Talk: Network Enabling Technology

TRINITY Long Room HUB

HMI Human Media Interaction

Speech Communication Laboratory

UNIVERSITÄT DES SAARLANDES

Social Signal Processing Network

Third Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisations in Speech

Dublin, October 26-27, 2012

Editors: Nick Campbell, Jürgen Trouvain, Khiet Truong
Cover design: Brian Vaughan
Proceedings: Ingmar Steiner

Programme

Friday, October 26, 2012

9.00– 9.15	Registration	
9.15– 9.30	Welcome Notes	
9.30–10.30	Invited speaker: Marc Mehu *The natural history of human laughter: evolution and social function*	1
10.30–11.00	Christelle Dodane, Jérémi Sauvage, Fabrice Hirsch & Melissa Barkat-Defradas *Is children's laughter related to their language development?*	2
11.00–11.30	Coffee and Tea Break	
11.30–12.00	Lesley Storey *Focus on laughter – a qualitative exploration of the social dimensions of laughter*	4
12.00–12.30	Mary Lavelle & Rose McCabe *What's so funny? An analysis of conversational laughter in schizophrenia*	6
12.30–13.00	Jennifer Hofmann, Willibald Ruch & Tracey Platt *The en- and decoding of schadenfreude laughter – Sheer joy expressed by a Duchenne laugh or emotional blend with distinct morphological expression?*	8
13.00–14.00	Lunch Break	
14.00–14.30	Radosław Niewiadomski, Sathish Pammi, Abhishek Sharma, Jennifer Hofmann, Tracey Platt, Richard Thomas Cruz & Bingqing Qu *Visual laughter synthesis: Initial approaches*	10
14.30–15.00	Jérôme Urbain, Hüseyin Cakmak & Thierry Dutoit *Development of HMM-based acoustic laughter synthesis*	12
15.00–15.30	Jocelynn Cu, Merlin Suarez & Madelene Sta. Maria *Towards building a context-based laughter model*	14
15.30–16.00	Coffee and Tea Break	
16.00–16.30	Sathish Pammi, Houssemeddine Khemiri & Gérard Chollet *Laughter detection using ALISP-based N-gram models*	16
16.30–17.00	Jérôme Urbain & Thierry Dutoit *Measuring instantaneous laughter intensity from acoustic features*	18
17.00–17.30	Anne van Leeuwen, Hugo Quené & Jos van Berkum *How do audible smiles and frowns affect speech comprehension?*	20
17.30–18.00	Rosemary Orr *Sorry, I just don't get it. . .*	22
18.00–19.00	Reception	

Saturday, October 27, 2012

Preface

Following the previous successful workshops on laughter held in Saarbrücken (2007) and Berlin (2009), we have the pleasure to welcome you to the third workshop of this series in Dublin.

The study of non-verbal vocal interaction is proving to be important to many research fields including human behaviour understanding, phonetics and discourse analysis, as well as more technology-oriented fields such as social signal processing and robotics. Previous research has shown that laughter and other nonverbal vocalisations (e.g., breath sounds, yawning, sighing) have important functions in social interaction, for example, providing feedback, signaling engagement, and regulating turn-taking. However, much about the phonetic characteristics of non-verbal vocalisations and the relationships between social functions and non-verbal vocalisation remains still unknown.

The goal of the present workshop is to bring together scientists from diverse research areas and to provide an exchange forum for interdisciplinary discussions in order to gain a better understanding of these processes. The workshop will consist of invited talks, oral presentations of ongoing research and discussion papers, and we are happy to include two excellent researchers as keynote speakers: Marc Mehu (Swiss Center for Affective Sciences) and Jens Edlund (KTH Stockholm).

The contributions to this meeting come from the fields of phonetics, linguistics, psychology, and human-machine interaction. The papers include many different aspects of laughter and non-verbal vocalisations related to research areas including multimodality, the 'timing together', affective states, social relations, and also computational models of non-verbal vocalisation.

Our warmest thanks go to the supporters who made this workshop possible: SSPnet, TCD Long Room Hub, TCD Speech Communication Lab, the SFI FastNet project at Trinity College Dublin, as well as Saarland University (Computational Linguistics and Phonetics) and the University of Twente (Human Media Interaction).

Nick Campbell, Jürgen Trouvain, and Khiet Truong

Dublin, Saarbrücken, and Enschede in October 2012

The natural history of human laughter: evolution and social function

Marc Mehu, Swiss Center for Affective Sciences, University of Geneva

Laughter is a peculiar human behaviour and has attracted the attention of researchers for decades. The occurrence of laughter worldwide suggests a long evolutionary history, with deep roots in the facial and vocal displays of primate ancestors. An ethological analysis of context and consequences suggests that laughter has most probably evolved as a social signal, which function is to create affiliative bonds among members of a species. Laughter, however, is also used in aggressive contexts, for example when undesired individuals are ostracized from a group. In both cases, the formation and maintenance of cooperative groups appears to be the essential feature through which laughter provides survival and reproductive benefits to individuals displaying it. I will review psychological and ethological evidence in favour of the hypothesis that laughter functions as a coalition formation signal. I will also stress the importance of gathering observational data on human laughter, as this line of evidence is necessary to complement experimental research on the production and perception of laughter. Finally I will discuss the implication of this research on the need, in social signal processing research, to study the social effects of emotional signals in addition to the symbolic meaning created by the community of perceivers.

Is children's laughter related to their language development?

Christelle Dodane*, Jérémi Sauvage*, Fabrice Hirsch** et Melissa Barkat-Defradas**

*Laboratoire DIPRALANG EA739, Université Montpellier 3
** Laboratoire PRAXILING UMR5237 CNRS & Université Montpellier 3

This paper aims at studying the acoustic development of child's laughter and its relation to language acquisition. To our knowledge, the first study dealing with the acoustic proprieties of laughter is due to Habermann (1955). In this work, the author provided anemographic data and defines laughter as a reflex-phenomenon with expiratory movements stopped by aspiratory pulses. Luchsinger and Arnold (1965) completed these findings using ultra-rapid imagery. They showed that laughter is characterized by a larynx in low position with expanded resonance cavities. As for its acoustic dimension, Trouvain (2003) defines laughter as a pattern of alternating non-voiced and voiced similar to a consonant-vowel syllable structure. Indeed, in the literature, laughter is often described along two segmental levels: a lower level, consisting of units that can be treated as consonants and vowels and a higher level that would be equivalent to the syllable (Bickley & Hunnicutt, 1992; Rothgänger *et al.*, 1998; Apte, 1983; Provine, 1993, 2003; Savithri, 2000). Unfortunately, very few studies were conducted on children's laughter (Nwokah et al., 1999; Mowrer, 1998; Tennis, 2009), while it first appears at about 4 months of age (Sroufe et Waters, 1976).

Our question can be formulated as follows: in laughter what is the part of physiology and what comes from linguistic development? We hypothesized that some acoustic characteristics of laughter are linked, on the one hand, with some physiological characteristics and, on the other hand, with speech properties. Thus, a study comparing laughter produced by congenitally deaf speakers and normally hearing speakers showed similar acoustical properties, which is consistent with the fact that laughter is established by human biology (Makagon et al., 1998). In terms of physiological changes, movements of the larynx, jaw and tongue are known to affect children's vocal repertoire, especially *f*o (Vihman, 1996; MacNeilage, 1998). Menard & Thibeault (2009) have shown that there were similarities and differences between children and adults as for prosodic representations. These differences in *f*o and amplitude of the voice can be linked to developmental movements of the larynx, glottal articulators being mastered earlier than supraglottal articulators. In terms of linguistic development, we know after Hallyday (1975) that syntax – which emerges gradually between 18 and 36 months – may affect the duration of laughter. The emergence of phonemes is also progressive and goes on, for French, until 48 months (Rondal, 1990). The production of certain segments in speech (such as open or closed vowel, oral or nasal sounds, occlusive or fricative consonants) may also be found in laughter. Therefore, we postulate that laughter segmental and supra segmental characteristics, measured trough the evolution of *f*o, the number of syllables and the nature of articulated sounds, will evolve along with physiological and linguistic development.

We used longitudinal data of three native French children (recorded from 18 to 36 months). Children were videotaped in natural interactions with their parents, so that their speech, moves and laughs were spontaneous and not artificially elicited in laboratory. Recordings are part of the Paris corpus, available on the CHILDES database. In order to establish the degree of linguistic development of each child, we measured their Mean Length of Utterances (M.L.U., Brown, 1973). The subject's fundamental frequency (*f*o) was calculated using an autocorrelation-based technique (Boersma, 1993). We performed acoustic analyses on 11 different parameters: total duration, proportion of voicing, fundamental frequency (mean *f*o, max. *f*o, min. *f*o, initial *f*o, final *f*o, max.-min. *f*o in semi-tones), number of syllables, number of phonemes and relative intensity. Then we determined the melodic type of each laugh. Statistical analyses were conducted with these 11 parameters, 3 subjects and 40 laughter series per subject (n= 120). We calculated ANOVAs for each dependant variable to determine if they could vary in function of several factors (i.e. age, M.L.U and/or subject).

Results showed that among the 11 acoustic parameters we investigated, only the relative intensity increased significantly with age ($F_{(1,118)}$=2,78, $p < .05$). This is probably due to the progressive maturation of the subglottic and glottic mechanisms. Indeed, during the development, *f*o and intensity are controlled before the supraglottic mechanisms (Vihman, 1996). Moreover, we noticed that laughs were produced with a large majority of rise-fall *f*o contours (60 % of all contours) and that their amount increases with age. This is in contradiction with the results of Savithri (2000) who showed that the most frequent *f*o contour in adults is a falling one. It seems that melodic configuration of laughter differs for children and adults. We also noticed a great inter-individual variability between children, which had already been observed in adults (Rothgänger et al., 1998). Ongoing analysis of the other parameters (i.e. duration of laughter, number of phonemes and formant structure of laughter) will show their (co)relation with MLU. We expect to observe an evolution of laughter as function of speech development. Statistical analysis and final results will be presented at the Workshop.

References

Apte, M.L. (1985). *Humor and Laughter. An Anthropological Approach*. Ithaca & London: Cornell University Press.

Bickley, C. & Hunnicutt, S. (1992). Acoustic analysis of laughter. *Proc. ICSLP Banff* (2), pages 927-930.

Boersma P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, 17, 97-110.

Brown, R. (1973). *A first language: The early stages*. Cambridge: MA : Harvard.

Habermann (1955), cit. in Luchsinger, R.L. & Arnold, G.E., *Voice, Speech and Language*, Constable and Co. Ltd.

Halliday, M.A.K (1975). *Learning how to mean ; explorations in the development of language*, E. Arnold, Londres.

Luchsinger, R.L. & Arnold, G.E. (1965). *Voice, Speech and Language*. Constable and Co., Ltd.

MacNeilage, P. (1998). The frame/content theory of evolution of speech production. *Behavior and Brain Sciences*, 21, 499-511.

Makagon et al. (2008). An acoustic analysis of laughter produced by congenitally deaf and normally hearing college students. *Journal of the Acoustical Society of America*, 124, 1, 47-483.

McKee, G., Malvern, D. & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literacy and Linguistic Computing*, vol. 15/3, 323-337.

Ménard L & Thibeault M. (2009). Développement de la parole et émergence de la structure prosodique chez l'enfant: une étude de l'accent d'emphase en français. *Canadian Journal of Linguistics* 54 (1), 117-136.

Mowrer, D.E. (1994). A case study of perceptual and acoustic features of an infant's first laugh utterances. *Humor*, 7 (2), pp. 139-155.

Nwokah, E.E., Hsu, H.-C., Davies, P. & Fogel, A. (1999). The integration of laughter and speech in vocal communication: a dynamic systems perspective. *Journal of Speech, Language & Hearing Research*, 42, pages 880-894.

Provine, R.R. (1993). Laughter punctuates speech: linguistic, social and gender contexts of laughter." *Ethology*, 95, pages 291-298.

Provine, R.R. (2003). *Le rire, sa vie, son œuvre*, Paris, Ed. Robert Laffont, 257 pages.

Rondal J.-A. (1990) *Votre enfant apprend à parler*. Bruxelles : Mardaga.

Rothgänger, H., Hauser, G., Cappellini, A.C. & Guidotti, A. (1998). Analysis of laughter and speech sounds in Italian and German students. *Naturwissenschaften*, 85, pages 394-402.

Savithri, S.R. (2000). Acoustics of Laughter. *Journal of Acoustical Society of India*, vol. 28, pages 233-238.

Sroufe, L.A. & Waters, E. (1976). The ontogenesis of smiling and laughter: A perspective on the organization of development in infancy. *Psychological Review*, 83, pages 173-189.

Tennis K. (2009). *The Acoustic Features of Children's Laughter*. Thèse soutenue à l'Université Vanderbilt (USA), 348 pages.

Trouvain, J. (2003). Segmenting phonetic units in laughter. *15th ICPhS*, Barcelona.

Vihman M. (1996). *Phonological development: the origins of languages in the child*. Cambridge, MA: Blackwell.

Focus on laughter: a qualitative study of the social dimensions of laughter

Author: Dr LHL Storey

Aim

This paper reports qualitative research carried out as part of the EU ILHILAIRE project looking at the way in which people talk about laughter and the social functions which it performs. The research provides a contribution to the ILHILAIRE project by looking at the social aspect of laughter from a qualitative perspective.

Method

Data was collected from focus groups of undergraduate students who received course credit for their involvement.

Focus Groups originated in the field of market research as a way of getting information about products for use in advertising. In academic research they are a data collection method which enable qualitative researchers to collect data in a way which incorporates interpersonal interaction between participants which provides an additional dimension unavailable in a 1-1 interview.

All participants were female undergraduate psychology students in the first year of their degree programme. The age range was from 18 to 55. The average group size was 5 participants. Each session lasted for just over an hour.

The focus group opened with a question about what made participants laugh and continued to explore positive and negative aspects of laughter, issues of social taboo in jokes and social bonding via laughter and joke telling.

Findings

The participants were universally positive about laughter and were able to discursively position problematic content eg racist jokes in a positive way. They explored the social functions of laughter in a variety of ways and again were reluctant to consider any possible negative consequences of (for example) “in-jokes”.

The interview schedule used for the focus group was open. The aims of the groups were to allow participants to discuss the positive and negative aspects of laughter without too many constraints. Participants were reluctant to consider that laughter had any negative aspects. When the issue of gelatophobia was introduced by the researcher it was met with incomprehension and disbelief and dismissed relatively quickly.

Discussion

The findings are discussed in relation to two main issues:

The role of gender in laughter and jokes

The use of laughter in emotional and social bonding.

These issues are discussed in relation to broader considerations for the use of laughter in social interactions.

Research funded by the ILHILAIRE EU project

What's so funny?
An analysis of conversational laughter in schizophrenia

Mary Lavelle & Rose McCabe

Background: Schizophrenia patients have difficulty interacting with others and are one of the most socially excluded groups in society.[1] The nature of patients' social exclusion is multifactorial. However, one contributing factor may be patients' behaviour during their social encounters. In a recent experimental study the undisclosed presence of a patient in a triadic interaction was found to change the nonverbal behaviour of patients' interacting partners.[2] Furthermore, patients' increased gesture use when speaking was associated with their partners experiencing poorer rapport with them.[2] This suggests that patients' partners may experience difficulty on an interpersonal level when interacting with a patient, which may contribute to patients' social exclusion.

Laughter can be as a marker of discomfort or awkwardness in social interaction.[3] In multiparty interaction, shared laughter may also indicate coalition between the laughing parties.[4, 5] This study investigated laughter in patients' triadic interactions with unfamiliar others, specifically focusing on laughter between patients' interacting partners as makers of interactional discomfort and coalition formation.

Method: The study consisted of two conditions: (i) a patient condition, comprising twenty patient groups (one schizophrenia outpatient and two healthy participants) and (ii) a control condition, comprising twenty control groups (three healthy participants). All interacting partners had not met prior to the study. Patients' partners were unaware of the patients' diagnosis and all participants were naive to the purposes of the study. Thus, the interactions were as naturalistic as possible within the motion capture environment. Interactions were audio-visually recorded using two, 2-D video cameras and simultaneously motion captured in 3-D (figure 1). Participants discussed a fictional moral dilemma called 'the balloon task', described elsewhere[2] and reached a joint decision on the outcome. Laughter was hand coded using the ELAN annotation tool. Patients' symptom severity was also assessed using the Positive and Negative Syndromes Scale.[6]

Preliminary results: Patients' partners displayed more shared laughter when patients had more positive symptoms such as hallucinations or delusional beliefs ($r(13)=.50$, $p=.04$). This was seen despite patients in the current study having only mild to moderate symptom levels and displaying no overt symptoms during the interaction.

A single case analysis of a patient's interaction was conducted using conversation analysis techniques. In this interaction, the shared laughter occurred after a lapse in the conversation. Specifically, at points in the interaction where the patient was expected to speak next but did not take the opportunity to do so. Shared laughter sometimes coincided with healthy participants' displaying mutual gaze and a sequence of highly synchronized nonverbal behaviours (e.g. one participant moving forwards as the other participant synchronously moves backwards) (figure 1).

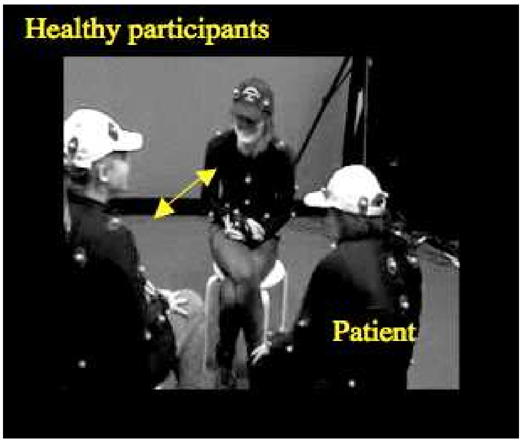

Figure 1. Healthy participants shared laughter

Discussion: These findings suggest that patients' partners displayed shared laughter, which the patient was not party to. This occurs more frequently when patients are more symptomatic. Preliminary analysis suggests that shared laughter occurs after lapses in the conversation where the patient is expected to speak but does not. Thus, the laughter of patients' partners may signal their shared interactional discomfort,[3,5] which may not be shared, by the patient.

This shared laughter suggests coalition formation between patients' partners.[4] Further qualitative analysis will be conducted to provide a more comprehensive understanding of the events prior to and during shared laughter in patients' interactions. This will also take into account the potential impact of disagreement on laughter patterns._ENREF_5[5]

The preliminary results of this study suggest that laughter may signal others' discomfort when a patient is not actively participating in the conversation. Moreover, laughter in multiparty interaction may be an indicator of coalition between pairs at specific points during social encounters. This may, in turn, influence participants' experiences of rapport in social interaction.

References

1. Social Exclusion Unit. Mental Health and Social Exclusion. In: Minister OotDP, ed. London; 2004.

2. Lavelle M, Healey PGT, McCabe R. Is nonverbal communication disrupted in interactions involving patients with schizophrenia? *Schizophrenia Bulletin* in press.

3. Haakana M. Laughter in medical interaction: From quantification to analysis, and back. *Journal of Sociolinguistics* 2002;6(2):207-235.

4. Bryant G. Shared laughter in conversation as a coalition signaling Paper presented at: XXI Biennial International Conference on Human Ethology 2012; Vienna, Austria.

5. Osvaldsson K. *On laughter and disagreement in multiparty assessment talk.* Vol 24. Berlin, ALLEMAGNE: Mouton de Gruyter; 2004.

6. Kay S, Friszbein A, Opler LA. The Positive and Negative Syndrome Scale for Schizophrenia. *Schizophrenia Bulletin* 1987;13:261-276.

Third Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisations in Speech Dublin, October 26-27, 2012

The en- and decoding of *schadenfreude* laughter

Sheer joy expressed by a Duchenne laugh or emotional blend with distinct morphological expression?

Jennifer Hofmann, Willibald Ruch, & Tracey Platt
Institute of Psychology Department of Personality and Assessment
University of Zurich
Zurich, Switzerland
j.hofmann@psychologie.uzh.ch

***Abstract*—This study investigates the facial features of *schadenfreude* laughter in historic illustrations by applying the Facial Action Coding System and assesses the decoding by naïve subjects. Results show that while the encoding of *schadenfreude* laughter is heterogeneous, *schadenfreude* is decoded when the facial expression unites markers of joy (Duchenne Display, consisting of the orbicularis oculi pars orbitalis muscle and the zygomatic major muscle), as well as markers of negative emotions (e.g., brow lowering), or in one case, where the initially categorized *schadenfreude* illustration contained markers distorting the expression of joy (e.g., frowning and the lowering of lip corners). These findings support the hypothesis that *schadenfreude* may be expressed by a morphologically distinct blend of a positive and a negative emotion, or is expressed by joyful laughter (with the expression being modulated due to social desirability).**

Keywords: schadenfreude, Facial Action Coding System, Duchenne Display

I. Introduction

Schadenfreude is an emotion often aligned with the expression of laughter [1,2]. The term expresses the pleasure derived when the envied [3,4], unliked [5] or dismissed [6] person or group experiences a mishap or embarrassing situation. It is expected that the strength of the emotion is moderated by the subjective perception of gratification [4,6,7].

Ekman [8] considers *schadenfreude* as one of 16 enjoyable emotions, which are facets of joy and expressed facially by the Duchenne Display (e.g., enjoyment smile or laugh, DD; [9]). Other authors claim *schadenfreude* to be a blend of joy and anger or taunt [2,10] or "malicious pleasure" [11,12]. Although not all languages have a specific word for this emotion (e.g., English), the existence of this specific feeling state also occurs in English and Anglo-American approaches to emotion classifications [8][1]. While *schadenfreude* should go along with a distinct emotional experience, it is unclear, whether its expression morphologically differs from joyful laughter (e.g., D-laughter, see [13,14,15]).

Three facial expressive patterns are feasible: Firstly, *schadenfreude* may be expressed by a D-laugh, as to the target, it is sheer joy seeing the "enemy" or out-group suffering. Secondly, *schadenfreude* may be expressed by a DD, but as deriving pleasure from the misfortunes of others is not socially desired (at least in Western cultures, see [8]), it might be that the laugh is regulated (e.g., down-regulation, modulation, or masking). This might involve elements of smile controls (i.e., lower intensity lip corner retraction and less opening of the mouth), additional covering actions like looking away, hand on mouth, or adding salient voluntary actions that change the expression of joy in the mouth and eye region. Thirdly, if *schadenfreude* is an emotion blend of a positive and negative emotion, it might be expressed facially by a blend of joy with markers of negative emotions (e.g., taunt).

Research on auditory features of *schadenfreude* laughter has been undertaken [2], but no current research team investigates the facial features of *schadenfreude*. Yet a recent review of the historic literature [16] shows that authors of the 19th and early 20th century [17,18,19,20,21], have made numerous attempts at distinguishing different qualities of laughter, with *schadenfreude* laughter being one of four converging categories (next to joyful laughter, intense laughter, and grinning) for which visual illustrations as well as verbal descriptions exist. Therefore, this study examines the en- and decoding of facial expressions of *schadenfreude* laughter in historic illustrations to see whether knowledge of historic authors can be a starting point for research on facial features of *schadenfreude* laughter.

II. Aims of the Study

The aims were twofold: firstly, all historic illustrations of types of laughter meeting a set of selection criteria (e.g., proposed laughter categories were available in visual illustration and verbal description and proposed by more than three illustrators; 18 illustrations in total; see Ruch, Hofmann, & Platt, under review) were coded by the Facial Action Coding System [22] by two independent certified coders to investigate the encoding of *schadenfreude* laughter (provided by Borée, Huter, and two illustrations by Rudolph). Secondly, the illustrations were presented in an online study to naive subjects (N = 87; 38 males; age M = 33.83, SD = 13.74; German-speaking background), which had to rate each illustration for their content of seven basic emotions (joy, anger, fear, sadness, surprise, disgust, contempt), as well as *schadenfreude*, shyness, maliciousness and friendliness on a five point answering scale.

---

[1] Words for *schadenfreude* are found outside the German language, e.g., in Chinese 幸灾乐祸, which means to gloat over someone else's misfortune or to take pleasure in others' misfortune (Pang, 2012; personal communication).

III. RESULTS AND DISCUSSION

Results of the encoding study show that there was no convergence in the Action Units (AU) encoded in the four *schadenfreude* illustrations and none of the illustrations entailed a pure DD. In fact, a small aperture of the mouth seemed to be characteristic (indicating a less forceful laughter exhalation or a down-regulated intensity due to display rules) and also the eye region showed unequal, usually lower intensities compared to the AU12 (lip corner puller). Most frequently an AU2 (outer brow raiser) could be observed in the eye region, as well as AU15 (lip corner depressor), AU17 (chin raiser) or AU20 (lip stretcher) in the mouth region. These actions help counteracting the effects of the contraction of the zygomatic major and orbicularis oculi muscles, i.e., distorting the expression of joy. This, in conjunction with the low(ered) intensity might help giving the appearance of that one is not really enjoying the mishaps of others in an unmitigated way. While this is compatible with the view that *schadenfreude* is a facet of joy [8], but down-regulated (lowered intensity) and concealed (additional AUs) due to lower cultural acceptance (in the encoder), it is doubtful whether or not decoders will identify it as such.

Results of the decoding study show that the four illustrations of *schadenfreude* differed in the extent to which they were perceived as containing *schadenfreude* (F [3,252] = 21.21, $p < .001$). Borée's illustration was rated to contain most *schadenfreude,* differing from all other illustrations, which did not differ from each other. This variant involves an AU7 (lid tightener) in extreme intensity, an AU1 and AU2 (inner and outer brow raiser), a very small mouth aperture and unilateral AU15 (lip corner depressor). However, this was also the only illustration where *schadenfreude* was rated at least "slightly present". Therefore, it was informative to compare all illustrations that received a mean rating of 2.00 or higher in *schadenfreude*, irrespective of the initial category (2 joyful, 1 intense, 1 *schadenfreude*). While these four generally did not differ (F [3,249] = 1.93, n.s.), there were convergences in their facial features. Three consisted of an AU4 (and/or AU9; brow lowerer, nose wrinkler) in addition to AU6 (cheek raiser), AU7 (lid tightener), and AU12 (lip corner puller), and an open mouth (AU25, AU26, AU27), plus the above described variant by Borée (1899) with the AU1, AU2 and AU15.

Those results support the claim that *schadenfreude* displays might entail the facial expression of enjoyment with strong mouth opening, plus a sign of negative emotion (AU4, AU9). This is further supported by the correlations of the emotion ratings for the four illustrations with high *schadenfreude* scores: *schadenfreude* ratings were highly correlated to rated joy ($r = .21$, $p < .05$), maliciousness ($r = .62$, $p < .001$), as well as contempt ($r = .40$, $p < .001$). Borée's variant supports the hypothesis of *schadenfreude* laughter being joy modulated by voluntary actions: a tightening of the eyelids, raising of eyebrows with asymmetric action in outer eyebrow and lips corner repression in addition to an AU12.

To conclude, while the encoding of *schadenfreude* was heterogeneous, decoding was linked to the appearance of markers of negative motions, namely AU4 and AU9. Next steps include the induction of *schadenfreude* in participants and the investigation of video clips portraying *schadenfreude* laughter. Clips of *schadenfreude* laughter of individuals feeling unobserved are needed to see whether the expression of *schadenfreude* consists of a pure joy laughter (D-laugh) when individuals feel unobserved and no social desirability leads to regulatory actions, or whether the occurrence of AU4 and AU9 can be replicated. Furthermore, the existence of *schadenfreude* should be investigated in different cultures to verify the universality of this feeling state.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°270780 (ILHAIRE).

REFERENCES

[1] Preuschoft, S. (1995). *"Laughter and smiling" in macaques – an evolutionary perspective.* Unpublished thesis. Utrecht: University of Utrecht.

[2] Szameitat, D. P., Alter, K., Szameitat, A. J., Wildgruber, D., Sterr, A., & Darwin, C. J. (2010). Acoustic profiles of distinct emotional expressions in laughter. *Journal of the Acoustical Society of America, 126,* 354-366.

[3] Smith, R. H., Turner, T. J., Garonzik, R., Leach, C. W., Urch-Druskat, V., & Weston, C. M. (1996). Envy and schadenfreude. *Personality and Social Psychology Bulletin, 22,* 158-168.

[4] Van Dijk, W. W., Ouwerkerk, J. W., Goslinga, S., & Nieweg, M. (2005). Deservingness and schadenfreude*. Cognition and Emotion, 19,* 933-939.

[5] Hareli, S. & Weiner, B. (2002). Dislike and envy as antecedents of pleasure at another's misfortune. *Motivation and Emotion, 26,* 257–277.

[6] Feather, N. T. & Sherman, R. (2002). Envy, resentment, schadenfreude, and sympathy: Reactions to deserved and undeserved achievement and subsequent failure. *Personality and Social Psychology Bulletin, 28,* 953-961.

[7] Feather, N. T. (2008). Effects of observer's own status on reactions to a high achiever's failure: Deservingness, resentment, schadenfreude, and sympathy. *Australian Journal of Psychology, 60,* 31-43.

[8] Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life.* New York: Times Books.

[9] Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology II. *Journal of Personality and Social Psychology, 58*, 342-353.

[10] Kemper, T. D. (1987). How many emotions are there? Wedding the social and the autonomic components. *The American Journal of Sociology, 93,* 263-289.

[11] Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

[12] Leach, C. W., Spears, R., Branscombe, N. R., & Doosje, B. (2003). Malicious pleasure: Schadenfreude at the suffering of another group. *Journal of Personality and Social Psychology, 84*, 932–943.

[13] Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology, 68,* 441-454.

[14] Keltner, D. & Bonanno, G. A. (1997). A study of laughter and dissociations: distinct correlates of laughter and smiling during bereavement. *Journal of Personality and Social Psychology, 73,* 687-702.

[15] Ruch, W. (1993). Exhilaration and humor. In M. Lewis & J. M. Haviland (Eds.), *The Handbook of Emotions* (pp. 605-616). New York: Guilford Publications.

[16] Huber, T. (2011). *Enkodierung und Dekodierung verschiedener Arten des Lachens. Eine FACS basierte Studie mit Schauspielern* [Encoding and decoding of different types of laughter. A FACS based study with actors]. Unpublished thesis, University of Zurich, Switzerland.

[17] Borée, A. (1899). *Physiognomische Studien* [Physiognomic studies]. Stuttgart: Julius Hoffmann.

[18] Heller, H. V. (1902). *Grundformen der Mimik des Antlitzes* [Basic forms of the mimic]. Vienna: Anton Schroll.

[19] Huter, C. (1925). *Physiognomik und Mimik* [Physiognomy and mimic]. Schwaig bei Nürnberg: Carl-Huter-Verlag.

[20] Piderit, Th. (1867). *Mimik und Physiognomik.* Detmold: Verlag der Meyerschen Hofbuchhandlung. [Original work published 1867]

[21] Rudolf, H. (1903). *Der Ausdruck der Gemütsbewegungen des Menschen* [The expression of the emotions in man]. Dresden: Gerhard Küthmann.

[22] Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System: A technique for the measurement of facial movement.* Palo Alto: Consulting Psychologists Press.

Radosław Niewiadomski[1], Sathish Pammi[1], Abhishek Sharma[1], Jennifer Hofmann[2], Tracey Platt[2], Richard Thomas Cruz[3], Bingqing Qu[1]

Visual laughter synthesis: Initial approaches

[1]*Telecom ParisTech, Rue Dareau, 37-39, 75014 Paris, France*
[2]*Universität Zürich, Binzmuhlestrasse, 14/7, 8050 Zurich, Switzerland*
[3]*De la Salle University, Manila, Philippines*

Visual laughter synthesis is a challenging task that was only rarely explored and empirical investigations are scarce. For the purpose of building a virtual agent able to laugh naturally we exploit different animation techniques such as a procedural animation or based on motion capture and we apply them to visual laughter synthesis. At the moment we focus on three approaches: procedural animation based on manual annotation of facial behavior; motion capture driven animation and animation generated from automatic facial movements detection. For the purpose of this study we use the Greta agent (Niewiadomski et al., 2011) that can be driven by both high-level anatomically inspired facial behavior description based on the Facial Action Coding System (FACS; Ekman, et al., 1978) or low-level facial animation parameterization (FAPs) that is a part of MPEG-4 standard (Ostermann, 2002) for facial animation. We also use two video corpora: AVLC database (Urbain et al., 2011) containing mocap, video and audio data of 24 subjects showing spontaneous amusement laugh responses and Queen's University Belfast's dataset of same sex interaction dyads during the watching of funny stimuli. We present all approaches in detail.

Manual annotation of action units. FACS is a comprehensive anatomically based system for measuring all visually discernible facial movement. It describes all distinguishable facial activity on the basis of 44 unique Action Units (AUs), as well as several categories for head and eye positions/movements and miscellaneous actions. Using FACS and viewing digital-recorded facial behavior at frame rate and in slow motion, certified FACS coders are able to distinguish and code all visually discernible facial expressions. Utilizing this technique, a selection of twenty pre-recorded, laboratory stimulated, laughter events were coded. These codes were then used to model the facial behavior on the Greta agent which is able to display any configuration of AUs. For 3 virtual characters single AUs were defined and validated by certified FACS coders under the constraints of the technology. A Behavior Markup Language (BML) implemented in Greta permits the control of each AU of the agent (its duration and intensity) independently. The animation of any AU is linearly interpolated according to Attack-Decay-Sustain-Release model (Ekman, et al., 1978). Next, the symbolic intensity values are converted to low-level facial animation parameters (FAP) which, finally, are used to deform a mesh of the virtual model. We also developed a tool that automatically converts manual annotation files created with Noldus Observer XT, a commercial tool for manual videos annotation, to BML. Consequently any file containing manual annotation of AUs can be easily displayed with the Greta agent.

Animation from automatic facial movements detection. The Greta agent uses facial animation parameters (FAPs) to realize low-level facial behavior. FAPs represent movements of MPEG-4 facial points compared to the 'neutral' face. In order to estimate FAPs of natural facial expressions, we made use of an open-source face-tracking tool – FaceTracker (Saragih et al., 2010) – to track facial landmark localizations. It uses a Constrained Local Model (CLM) fitting approach that includes a Regularized Landmark Mean-Shift (RLMS) optimization strategy. It can detect 66 facial landmark coordinates within real-time latency depending on the system's configuration.

Facial geometry differs from one human to another one. Therefore, it is difficult to estimate FAPs without neutral face calibration. To compute FAPS from facial landmarks, a neutral face model is created with the help of 50 neutral faces of different persons. With the help of this model, FAPs are estimated as the distance between facial landmarks and neutral face landmarks. In case of user-specific FAP estimation in a real-time scenario, the neutral face is estimated from a few seconds of video by explicitly requesting the user to hold the face still. However, the better estimation of FAPs requires manual intervention for tweaking weights to map landmarks and FAPs, which is a downside of this methodology.

The landmark coordinates produced by the FaceTracker are observed as noisy due to the discontinuities and outliers in each facial point localization. Especially, the realized behavior is unnatural on a virtual model when we re-target the observed behavior onto the Greta agent. To smooth the face-tracking parameters, a temporal regression strategy has been applied on individual landmarks by fitting 3rd order polynomials using a sliding window, where the sliding window size and its shifting rate are 0.67 seconds and 0.33 seconds respectively.

Animation from motion capture data. AVLC corpus (Urbain et al., 2011) contains motion capture data of laugh episodes that have to be retargeted to the virtual model. The main problem in these kinds of approaches consists in finding appropriate mappings for each participant's face geometry and different virtual models. Many existing solutions are typically linear (e.g., methods based on blend shape mapping) and do not take into account dynamical aspects of the facial motion itself. Recently, Matthew Zeiler and colleagues (2011) proposed to apply variants of Temporal Restricted Boltzmann Machines (TRBM) to the facial retargeting problem. TRBM are a family of models that permit tractable inference and allows complicated structures to be extracted from time series data. These models can encode a complex nonlinear mapping from the motion of one individual to another, which captures facial geometry and dynamics of both source and target. In the original application (Zeiler et al., 2011) these models were trained on a dataset of facial motion capture data of two subjects, asked to perform a set of isolated facial movements based on FACS. The first subject had 313 markers (939 dimensions per frame) and the second subject had 332 markers (996 dimensions per frame). Interestingly there was no correspondence between marker sets. They were able to retarget the motion with a RMS error of 2 %. However, they only evaluated their results on slow facial movements.

We use TRBM models for our project, which involves retargeting from an individual to a virtual character. In our case, we take the input as the AVLC mocap data and output the corresponding facial animation parameters (FAP) values. This task has two interesting aspects. First, the model performance was previously evaluated only on retargeting an isolated slow expression whereas our case involves transitions from laughter to some other expression (smile or neutral) as well as very fast movements. Second, we use less markers compared to the original application. Our mocap data had only 27 markers on face, which is very sparse.

So far we used the AVLC data of one participant. As a training set we used two sequences, one of 250 frames and another one of 150 frames. Target data (i.e., facial animation parameters) for this training set was generated using manual retargeting procedures explained in Urbain et al. (2011). Both the input and output data vectors were reduced to 32 dimensions by retaining only their first 32 principal components. Since this model typically learns much better on scaled data (around [-1,1]), the data was then normalized to have zero mean and scaled by the average standard deviation of all the elements in the training set. Having trained the model, we used it to generate facial animation parameters values for 2 minutes long mocap data (2500 frames coming from the same participant). The first results are promising but more variability in the training set is needed to retarget more precisely different type of movements.

Conclusion. These three approaches offer different degrees of flexibility and control over the expression, different levels of realism and precision of the movements. We expect, for instance, that the mocap-based animation should be richer in movements and consequently it may be perceived as more realistic. Also using mocap data should permit to maintain the temporal and dynamic characteristics of the original laugh. On the other hand animation generated with this method is difficult to control manually (e.g., its duration, intensity, communicative function). Moreover the mocap procedure is invasive, recourse- and time consuming. On the other hand, describing animation by action units allows one to control precisely an animation and its meaning (e.g., by adding or removing AU6, a marker of the Duchenne smile) but has all the weaknesses of procedural approaches to facial animation. The animation is poor in details and the dynamics of the movements is very simplistic. Finally, a solution based on the automatic facial action detection combines advantages of both solutions: it should be sufficiently rich in the details (it depends highly on the quality of the face tracker applied). At the same time one can manually control and edit the final animation by adding or removing some facial actions. Still it requires that recordings be taken in controlled conditions (e.g., good lighting).

Future works will consist of a set of perceptive studies that we want to develop in order to check the quality of the animations and compare our 3 methods. For this purpose we use just one set of laugh episodes and will generate animations with these 3 different approaches. The factors considered in the evaluation will be believability and naturalness of the animations.

Bibliography

Ekman, P., Friesen, W.V., & Hager, J. C. (1978). *Facial Action Coding System: A technique for the measurement of facial movement*. Palo Alto: Consulting Psychologists Press.

Niewiadomski, R., Bevacqua, E., Quoc Anh Le, Obaid, M., Looser, J., & Pelachaud, C. (2011). Cross-media agent platform. *Web3D ACM Conference*, Paris, France (pp. 11-19).

Ostermann, J. (2002). Face animation in MPEG-4. In I. Pandzic and R. Forchheimer (eds.), *MPEG-4 Facial Animation - The Standard Implementation and Applications* (pp. 17–55), England: Wiley.

Saragih, J. M., Lucey, S., & Cohn, J. F. (2011). Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision 91* 200 – 215.

Urbain, J., Niewiadomski, R., Bevacqua, E., Dutoit, T., Moinet, A., Pelachaud, C., Picart, B., Tilmanne, J., & Wagner, J. (2010). AVLaughterCycle. Enabling a virtual agent to join in laughing with a conversational partner using a similarity-driven audiovisual laughter animation. *Journal of Multimodal User Interfaces, 4,* 47-58.

Zeiler, M.D., Taylor, G.W., Sigal, L., Matthews, I., & Fergus, R. (2011). Facial Expression Transfer with Input Output Temporal Restricted Boltzmann Machines. *Neural Information Processing Systems Conference NIPS 2011*, Granada, Spain. (pp. 1629-1637).

Third Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisations in Speech Dublin, October 26-27, 2012

Development of HMM-based acoustic laughter synthesis

Jérôme Urbain*, Hüseyin Cakmak, and Thierry Dutoit

TCTS Lab, Faculty of Engineering, University of Mons, Belgium

Laughter is a key signal in human communication, conveying information about our emotional state but also providing social feedback to the conversational partners. With the development of more and more natural human-computer interactions (with the help of embodied conversational agents, etc.), the need emerged to enable computers to understand and express emotions. In particular, to enhance human-computer interactions, talking machines should be able to laugh.

Yet, compared to speech synthesis, acoustic laughter synthesis is an almost unexplored domain. Sundaram and Narayanan [5] modeled the laughter intensity rhythmic envelope with the equations governing an oscillating mass-spring and synthesized laughter vowels by Linear Prediction. This approach to laughter synthesis was interesting, but the produced laughs were judged as non-natural by listeners. Lasarcyk and Trouvain [3] compared laughs synthesized by an articulatory system (a 3D modeling of the vocal tract) and diphone concatenation. The articulatory system gave better results, but they were still evaluated as significantly less natural than human laughs.

To improve laughter synthesis naturalness, we propose to use Hidden Markov Models (HMMs), which have proven efficient for speech synthesis. We opted for the HMM-based Speech Synthesis System (HTS) [4], as it is free and widely used in speech synthesis and research. The data used comes from the AVLaughterCycle database (AVLC), which contains around 1000 laughs from 24 subjects and includes phonetic transcriptions of the laughs [6].

HTS provides a demonstration canvas for speech synthesis, which enables to quickly obtain synthesis models with standard speech parameters. Our first works were to use this canvas to build a baseline for HMM-based laughter. Then, we looked at adapting our data and modifying some parts of the HTS demo to improve the quality of the obtained laughs.

The major improvement of the AVLC database to better exploit the potential of HTS is the annotation of laughter "syllables"[1]. This enables to include contextual parameters (e.g. the position of the "phoneme" within its "syllable", the position of the current "syllable" within the current "word", etc.) in the synthesis models.

*jerome.urbain@umons.ac.be

[1]We use quotation marks around the terms *syllable, phoneme and word* to distinguish the laughter units from their speech counterparts.

Two important modifications have also been done in the HTS process compared to the demonstration algorithms. First, the standard Dirac pulse train for voiced excitation has been replaced by the DSM model [1], which better fits the human vocal excitation shapes and reduces the buzziness of the synthesized voice. Second, the standard vocal tract and fundamental frequency estimation algorithms provided by HTS have been replaced by the STRAIGHT method [2], which is known in speech processing to provide better estimations.

These modifications largely improved the quality of the synthesized laughs. Some examples of HMM-based laughter synthesis are available on `http://www.ilhaire.eu/blog~Acoustic-Laughter-Synthesis`. It is important to note that we are currently not able to generate new laughter phonetic transcriptions, and in consequence we re-synthesize existing human transcriptions. Future work includes the development of a module to generate (or modify existing) phonetic transcriptions, further optimizations of the synthesis parameters and a perceptive evaluation study to quantify the improvements and provide a benchmark for future developments.

Acknowledgment

This work was supported by the European FP7-ICT-FET project ILHAIRE (grant $n°$270780).

References

[1] T. Drugman and T. Dutoit. The deterministic plus stochastic model of the residual signal and its applications. *Audio, Speech, and Language Processing, IEEE Transactions on*, 20:968–981, 2012.

[2] H. Kawahara. Straight, exploitation of the other aspect of vocoder: Perceptually isomorphic decomposition of speech sounds. *Acoustical science and technology*, 27(6):349–353, 2006.

[3] E. Lasarcyk and J. Trouvain. Imitating conversational laughter with an articulatory speech synthesis. In *Proceedings of the Interdisciplinary Workshop on the Phonetics of Laughter*, pages 43–48, Saarbrücken, Germany, August 2007.

[4] Keiichiro Oura. Hmm-based speech synthesis system (hts) [computer program webpage]. `http://hts.sp.nitech.ac.jp/`, consulted on June 22, 2011.

[5] S. Sundaram and S. Narayanan. Automatic acoustic synthesis of human-like laughter. *Journal of the Acoustical Society of America*, 121(1):527–535, January 2007.

[6] Jérôme Urbain and Thierry Dutoit. A phonetic analysis of natural laughter, for use in automatic laughter processing systems. In *Proceedings of the fourth bi-annual International Conference of the HUMAINE Association on Affective Computing and Intelligent Interaction (ACII2011)*, pages 397–406, Memphis, Tennesse, October 2011.

Towards building a context-based laughter model

Jocelynn Cu, Merlin Suarez, Madelene Sta. Maria
Center for Empathic Human-Computer Interactions
De La Salle University

Abstract

Laughter is a significant social cue that contributes to a meaningful interaction. It is a known regulatory mechanism in the expression of emotion (Mesquita and Frijda, 1992). As such, we would like to answer the questions what are the different types of emotions expressed through laughter, and how can these laughter be distinguished from each other?

In our previous study on laughter (Suarez et. al.), we built a multimodal laughter corpus to study how different emotions are expressed through laughter. As in Urbain et. al. (2010), we gather acted data from 2 professional actors (one male and one female). The actors were asked to express five emotions through laughter (rough English equivalents are given, but in actual data collection, emotion label is given in the local language to dispel semantic misinterpretation) and these are *natutuwa* (happiness), *kinikilig* (giddiness), *nasasabik* (excitement), *nahihiya* (embarrassment), and *mapanakit* (hurtful). According to local linguists, these five labels typically describe the emotions carried by laughter in natural interactions. Each actor was interviewed at the end of each enactment to explain the motivation for their expression. These served as contextual information which psychologists relied on to label the emotion. We also collected induced laughter (as in Nachami and Santhanam, 2008) from three subjects who were watching funny videos or comic strips. In this case, subjects were asked to label their own emotions using the five labels and using the valence-arousal dimensions through FeelTrace. Volunteer annotators who have a high empathy quotient score also provided discrete and dimensional labels. From these clips, we extract audio features and facial features to build laughter models.

Based on our findings, we learned that we can better distinguish the different types of emotions in laughter though the audio information rather than the facial information. However, these models were not robust. The mapanakit (hurtful) laughter was not elicited properly and it was difficult to properly distinguish natutuwa (happiness), kinikilig (giddiness), and nasasabik (excitement) laughters from each other . We also found that contextual information is significant to correctly identify the type of emotion expressed in the laugh. This concern was evident when the psychologists were annotating the acted data and when the volunteers were annotating the induced data.

Although these are acted and induced data, we notice that there is a masking phenomenon that occurs when the subject is laughing. We label this as restrained laughter. In restrained laughter, the person is holding back his expression of laughter because he/she could be hiding something or suppressing the expression of another emotion. Whatever this emotion is, we cannot determine because our basis is the audio and face only. However, given the context and background information why this person is laughing in a particular way helps us determine this specific emotion. We believe that this phenomenon can be related to cultural influences and display rules. Greater suppression is found to happen among those who live in cultures that highly endorse the expression of positive emotions (Matsumoto et. al., 2008; Safdar et. al., 2009).

To understand the difference between restrained and unrestrained laughter, we plan to build a model that will consider contextual information when interpreting the emotions expressed through laughter. To do this, we need to study laughter in a more natural setting. We will collect laughter samples of subjects engaged in a spontaneous interaction with another subject or with a group of people. However, the spontaneous interactions will have to be done in a controlled environment where we can still capture clear audio and video from the subjects. We will annotate the clips with additional information to form the context, which includes the following: the type of interaction (dyadic or multiparty interaction), profile (age and gender of the subject), role of the speaker (leader or peer), relationship of the subject to the other (friends, co-workers, boss, etc.), topics being discussed (family, work, politics, relationships, etc.), number of times the subject is laughing, description of laughter (duration, intensity, volume), direction of laughter (laughter at self or at others), purpose of laughter (to convey emotion, to punctuate a statement, to change the social atmosphere, etc.), collective mood of the interaction, type of laughter (voiced/unvoiced, restrained/unrestrained, overlapping/sequential, etc.).

With these contextual information, we will investigate the use of appraisal theory in building the context-based laughter model. There were several existing appraisal models (Becker-Asano, 2008; Marsella and Gratch, 2009;

Marinier, Laird, and Lewis, 2009) that has been successful in incorporating contextual information to determine a person's affect. We will use knowledge-driven rules and data-centric probabilities to map contextual information into discrete laughter emotions. Specifically, seven appraisal variables introduced by Scherer will be used to derive the laughter emotion, which are suddenness, intrinsic pleasantness, goal relevance, unpredictability, outcome probability, discrepancy from expectation, and goal conduciveness. The model will be evaluated based on the annotated spontaneous laughter clip collection.

References

Becker-Asano, C. (2008) Wasabi: Affect simulation for agents with believable interactivity. Unpublished doctoral dissertation, Faculty of Technology, University of Bielefeld. (IOS Press (DISKI 319))

Marsella, S. C. and Gratch, J. (2009) EMA: A process model of appraisal dynamics. Journal of Cognitive Systems Research. 10(1):70–90.

Marinier, R. P. III, and Laird, J. E. (2008) A computational unification of cognitive behavior and emotion. Journal of Cognitive Systems Research.

Matsumoto, D., Yoo, S-H., Fontaine, J., Anguas-Wong, A. M., Arriola, M.,Ataca, B., et al. (2008). Mapping expressive differences around the world: The relationship between emotional display rules and individualism vs. collectivism. *Journal of Cross-Cultural Psychology, 39,* 55–74.

Mesquita, B. & Frijda, N.H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, 112: 179-204.

Nachami, M. and Santhanan, T. (2008) Laughter inquisition in affect recognition. Journal of Theoretical and Applied Information Technology, 429 – 432.

Safdar, S., Friedlmeier, W., Matsumoto, D.; Yoo S-H., Kwantes, C., Kakai, H. et al. (2009). Variations of emotional display rules within and across cultures: A comparison between Canada, USA, and Japan. *Canadian Journal of Behavioural Science*, 41: 1-10.

Suarez, M., Cu, J., Sta. Maria, M. (2012) Building a multimodal laughter database for emotion recognition. LREC 2012, Istanbul, Turkey.

Szameitat, D. P., Szameitat, A. J., Wildgruber, D., Dietrich, S., Alter, K., Darwin, C. J., and Sterr, A. (2009) Differentiation of Emotions in Laughter at the Behavioral Level. Emotion 9(3): 397 – 405.

Urbain, J., Bevacqua, R., Dutoit, T., Moinet, A., Niewiadomski, R., Pelachaud, C., Picart, B., Tilmanne, J., and Wagner, J. (2010) The AVLaughterCycle Database. LREC 2010, 2996 – 3001.

Laughter detection using ALISP-based N-Gram models

Sathish Pammi, Houssemeddine Khemiri and Gérard Chollet
Telecom ParisTech, Rue Dareau, 37-39, 75014 Paris, France
{firstname.lastname}@telecom-paristech.fr

Laughter is a very complex behavior that communicates a wide range of messages with different meanings. It is highly dependent on social and interpersonal attributes. Most of the previous works (e.g. [1, 2]) on automatic laughter detection from audio uses frame-level acoustic features as parameters to train their machine learning techniques, such as Gaussian Mixture Models (GMMs), Support Vector Machines (SVMs) etc. However, segmental approaches that capture higher-level events have not been adequately focussed due to the nonlinguistic nature of laughter. This paper is an attempt to detect laughter regions with the help of automatically acquired acoustic segments using Automatic Language Independent Speech Processing (ALISP) [3, 4] models.

Method

The ALISP tools provide a general framework for creating speech units with little or no supervision. As shown in the Figure 1, the ALISP models are estimated on an audio database through parametrization, temporal decomposition, vector quantization, and Hidden Markov Modeling (HMM). ALISP units/segments are automatically acquired (i.e. unsupervised) segmental units from the ALISP models.

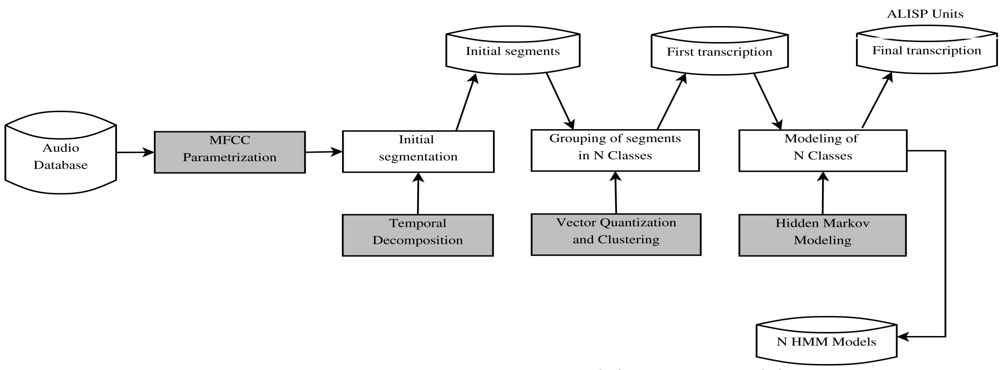

Figure 1: ALISP units acquisition and their HMM modeling

This work uses an ALISP-based automatic segmentation system which is modeled with 26 days of complete broadcast audio of 13 French radio stations provided by YACAST. This model can be considered as an universal acoustic model because of its training database includes all possible sounds like music, laughter, advertisements etc. The advantage of these models is not only the capability of segmenting any audio, but also providing appropriate symbolic level annotation for the segments. In order to represent ALISP units, the segmentation system uses 64 ALISP symbols (such as *'Ha', 'Hv'* and *'H@'*) in addition to a silence label. Figure 2 is an example of laughter audio segmented by ALISP models.

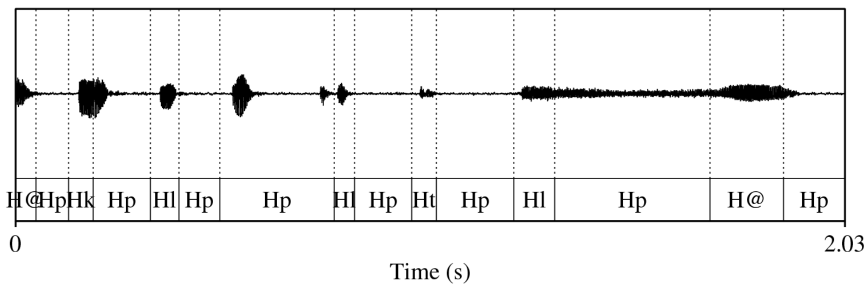

Figure 2: Laughter audio segmented by ALISP models

We hypothesize that the sequence of ALISP symbols contains the patterns of laughter. N-gram models (e.g. the sequence 'Hp-H@-Hp' is a 3-gram) on the ALISP symbolic sequence could model the patterns to detect laughter regions. A tool has been built to detect laughter from audio using linearly combined estimate of N-gram models of increasing order as follows:

$$\hat{P}(w_n \mid w_{n-2,} w_{n-1}) = \lambda_1 P(w_n \mid w_{n-2,} w_{n-1}) + \lambda_2 P(w_n \mid w_{n-1}) + \lambda_3 P(w_n)$$

$$\text{Where: } \sum_i \lambda_i = 1$$

In the above equation, trigram (i.e. N=3) models are mixed with bigram and unigram models. The linear interpolation of N-gram models ensure that the models suffer less from sparseness.

Evaluation and results

The ALISP-based N-gram models are trained on SEMAINE-DB [5] and AVLaughterCycle [6] databases, and the models are evaluated with Mahnob laughter database [7]. All of the three databases have manual annotations of laughter cycles. The SEMAINE-DB contains 5015 and 389 seconds of non-laughter and conversational laughter audio respectively; whereas the AVLaughterCycle DB has 3477 seconds of hilarious laughter. The MAHNOB laughter database contains 1837 and 2307 seconds of laughter and non-laughter audio respectively. As shown in Figure 3, we compared ALISP-based N-Gram models with acoustic models like GMMs, sequential (left-to-right) HMMs and ergodic (fully-connected) HMMs trained to discriminate laughter and non-laughter audio. Simple GMMs performed better precision when compared to HMMs, while ergodic HMMs provides high recall rate (93%) than GMMs. ALISP-based N-Gram models have good precision in detecting laughter, though, the recall rate is low. For example, the interpolated 5-Gram ALISP model showed more than 90% precision which indicates minimum manual intervention to find false alarms while extracting laughter from naturalistic audio resources such as radio broadcasting.

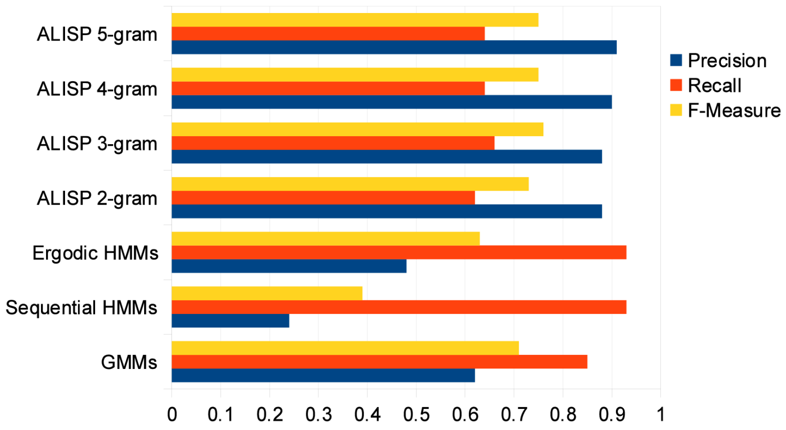

Figure 3: Performance of ALISP-based N-gram models versus GMM and HMM-based acoustic models

Discussion

The performance of ALISP-based N-grams models can be improved with more laughter training material. The ALISP symbols could be assumed as descriptions of 'very short acoustic acts'. The sequence of such cues could preserve the behavioral patterns of not only laughter, but also any other interactional vocalizations that are nonlinguistic in nature. We plan to investigate possibilities combine frame-level acoustic features with segmental features to improve the performance of laughter detection.

Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 270780 (ILHAIRE project). We would like to thank Prof. Dijana Petrovska for useful discussions.

References

[1] Truong, K.P. and Van Leeuwen, D.A. (2007), “Automatic discrimination between laughter and speech”, journal of Speech Communication.
[2] Knox, M. and Mirghafori, N. (2007), “Automatic laughter detection using neural networks”, proceedings of INTERSPEECH 2007.
[3] Chollet G, Cernocký J, Constantinescu A, Deligne S, Bimbot F (1999), “Towards ALISP: a proposal for Automatic Language Independent Speech Processing”, NATO ASI Series. Springer, pp 375-387.
[4] Khemiri H, and Chollet G, and Petrovska-Delacrétaz D (2011), “Automatic detection of known advertisements in radio broadcast with data-driven ALISP transcriptions”, 9th International Workshop on Content-Based Multimedia Indexing (CBMI), pp 223 – 228.
[5] McKeown, G. and Valstar, M.F. and Cowie, R. and Pantic, M. (2010), “The SEMAINE corpus of emotionally coloured character interactions”, IEEE International Conference on Multimedia and Expo (ICME), pp 1079–1084.
[6] Urbain, J. and Niewiadomski, R. and Bevacqua, E. and Dutoit, T. and Moinet, A. and Pelachaud, C. and Picart, B. and Tilmanne, J. and Wagner, J., (2010) “AVLaughterCycle: Enabling a virtual agent to join in laughing with a conversational partner using a similarity-driven audiovisual laughter animation”, Journal on Multimodal User Interfaces, pp. 47 – 58.
[7] Petridis S, Martinez B, Pantic M (nd.), “MAHNOB-Laughter Database”, Retrieved from: http://ibug.doc.ic.ac.uk/research/mahnob- laughter-database.

Measuring instantaneous laughter intensity from acoustic features

Jérôme Urbain* and Thierry Dutoit

TCTS Lab, Faculty of Engineering, University of Mons, Belgium

Being able to process and express emotional signals when interacting with humans is an important feature for machines acting in roles like companions or tutors. Laughter is a very important signal that regulates human conversations. It is however hard for machines, which usually have no real comprehension of the phenomenon that triggered laughter, to understand the meaning of human laughs and, in consequence, to react accordingly.

In this paper, we explore one dimension to characterize laughs: their intensity[1]. Without better understanding the conversation, a machine that can infer the intensity of users' laughs will be better equipped to select an appropriate answer (which can be laughing at an intensity related to the detected laugh).

In [1], an online evaluation study has been conducted where naive raters where asked to estimate the intensity of audiovisual laughter clips. They had to provide one intensity value for each laugh, on a 5-point Likert scale ranging from very low intensity to very high intensity. Each episode of the AVlaughterCycle database [2], that contains one thousand laughs recorded with a webcam and a head-mounted microphone, has been rated by at least 6 different participants. Audiovisual features that correlate with the global intensity have been identified. For example, the range of MFCC0 (related to the acoustic energy) and the maximum opening of the mouth over a laughter episode are correlated with the perceived intensity of this episode.

Here, we extend this work by investigating the possibility to automatically draw instantaneous intensity curves. The advantages compared to the global intensity value are the following: 1) the intensity can be obtained in real-time, without waiting for the end of the episode 2) such curves can be useful to drive laughter synthesis 3) instantaneous intensity curves can provide more insight to understand what creates the local and global perceptions of intensity.

To do so, 49 laughs uttered by 3 speakers of the AVLaughterCycle database and ranging across the global intensity values scored in [1] have been manually annotated by one rater. One intensity value was assigned every 10ms, using only the audio signal.

A linear combination of acoustic features has been designed to match the manual annotation. Figure 1 displays the manual intensity for one laugh, together with the intensity predicted from 2 audio features: loudness (i.e. the perceived acoustic amplitude) and F0. The automatic intensity curve is a weighted sum of these two features, followed by median filtering to smooth the output.

*jerome.urbain@umons.ac.be

[1] In this paper, the term *intensity* is used to refer to how amused the laugher seem to be.

We can see that the automatic curve is matching the trend of the manual annotation. Furthermore, the overall laughter intensity can be extracted from the continuous annotation curve: correlation coefficients between the median intensity scored by users and the intensity predicted from acoustic features are over 0.7 for 21 out of 23 subjects[2].

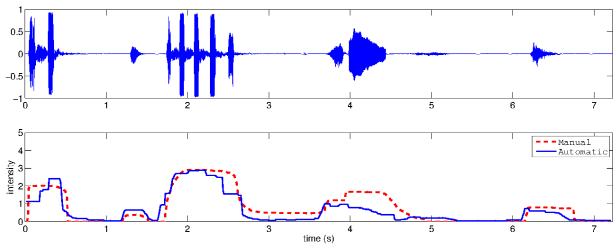

Figure 1: Example of laughter continuous intensity curve. Top: waveform; Bottom: manual and automatic intensity curves.)

Work is in progress to optimize the computation of the continuous intensity (with a trained algorithm using several features instead of the manually designed linear combination) and ensure that the intensity values fall in the same ranges for all the subjects. This is indeed one issue of the weighted sum: it is able to detect which laugh or laughter segment is more intense than another one within one subject (which explains the high correlations given above), but the range of values differ from one participant to the other. To overcome these problems, we are currently training neural networks. The first results are promising both for the continuous and global intensities, but we still have to smooth the continuous curves and perform cross-validation. In addition, we could consider adding visual features to have more robust estimations (in particular for laughs with low acoustic contributions).

Acknowledgment

This work was supported by the European FP7-ICT-FET project ILHAIRE (grant n°270780).

References

[1] R. Niewiadomski, J. Urbain, C. Pelachaud, and T. Dutoit. Finding out the audio and visual features that influence the perception of laughter intensity and differ in inhalation and exhalation phases. In *Proc of the ES^3 Workshop, Satellite of LREC 2012*, Istanbul, Turkey, May 2012.

[2] Jérôme Urbain, Elisabetta Bevacqua, Thierry Dutoit, Alexis Moinet, Radoslaw Niewiadomski, Catherine Pelachaud, Benjamin Picart, Joëlle Tilmanne, and Johannes Wagner. The AVLaughterCycle database. In *Proc. of LREC'10*, Valletta, Malta, May 2010.

[2] The 24th subject of the AVLaughterCycle database only laughed 4 times, each time with an intensity 1, which prevents us from computing correlations

How do audible smiles and frowns affect speech comprehension?

Anne van Leeuwen MA, prof. Dr. Hugo Quené and prof. Dr. Jos van Berkum
Utrecht Institute of Linguistics OTS, Utrecht University

We often smile (and frown) while we talk. Listeners to such affective speech have to integrate the affective and the linguistic cues in the speech signal. Following up on earlier work (see Quené et al., 2012), we investigated whether and how affective phonetics (i.e. vocal expressions such as smiling) and affective semantics (sentence-level meaning) interact during spoken language comprehension of sentences and how *perspective* modifies these interactions. We explored this by presenting phonetically and semantically manipulated spoken Dutch sentences to listeners while collecting behavioral and neural (ERP) measurements.

The target materials consisted of utterances that contained a positive or negative content word. Additionally, perspective was taken into account so that sentences were in first person ('ik') or in third person ('hij' or 'ze'). Utterances were phonetically manipulated using Praat's LPC analysis and resynthesis. Between analysis and resynthesis, the formant frequencies were manipulated (upwards or downwards shift: 10%) to imitate the spectral effects of facial expressions while talking (Ohala, 1980; Quené et al., 2012). This resulted in affective congruent realizations (positive – smiling, negative – frowning), or affective incongruent realizations (negative – smiling, positive – frowning). As a control measure we constructed neutral utterances that were as similar as possible to the target materials, but without being affective in nature. These sentences had a low frequent or high frequent content word and they carried neutral articulation. The same LPC analysis and resynthesis were used, but without altering the formants. This resulted in phonetically neutral synthetic realizations (high frequent – neutral, low frequent neutral). See examples below.

Table 1: Example sentences

☺ = smile manipulation
☹ = frown manipulation
😐 = neutral

EXPERIMENTAL DESIGN[1]	Perspective	
	1st person	**3rd person**
Congruent	Ik word erg *vrolijk* van hem ☺ Ik word erg *somber* van hem ☹	Hij/ze wordt erg *vrolijk* van hem ☺ Hij/ze wordt erg *somber* van hem ☹
Incongruent	Ik word erg *somber* van hem ☺ Ik word erg *vrolijk* van hem ☹	Hij/ze wordt erg *vrolijk* van hem ☹ Hij/ze wordt erg *somber* van hem ☺
Low frequent	Ik word erg *bedrijvig* hierdoor 😐	Hij/ze wordt erg *bedrijvig* hierdoor 😐
High frequent	Ik word erg *wakker* van koffie 😐	Hij/ze wordt erg *wakker* van koffie 😐

In the EEG study, participants just listened to the utterances while we measured their brain response 200 ms before until 1000 ms after the onset of the critical word. In the behavioral study, two different groups of participants were asked to jugde wether the utterance (truncated at the offset of the critical word) was

[1] Gloss:
- Ik/hij/ze word(t) erg vrolijk/somber van hem
  'I/he/she become(s) very cheerful/sad because of him'
- Ik/hij/ze word(t) erg bedrijvig/wakker …
  'I/he/she become(s) very active/alert …

positive/negative (in terms of meaning), or smiling/frowning (in terms of articulation).

The general predictions were that incongruent sentences were responded to more slowly and eliciting a greater N400 component than congruent sentences. We were especially interested in the effect of perspective on affective sentences: would listening to first person perspective result in a qualitative and/or quantitative different response than the responses to third person perspective? We hypothesized that perspective would modify the effects found in the affective sentences, but not in the control sentences because in the first case first person perspective conveys direct information about the affective state of the speaker, while in the latter case there is no affective information in the speech signal.

Results will be discussed in the light of affective processing and how smiling and frowning influence the way listeners process speech. We will also discuss perspective and how perspective modifies responses to affective stimuli.

References

(1) Ditman, T., Brunye, T. T., Mahoney, C. R., & Taylor, H. A. (2010). Simulating an enactment effect: Pronouns guide action simulation during narrative comprehension. *Cognition, 115*(1), 172-178.

(2) Drahota, A., Costall, A., & Reddy, V. (2008). The vocal communication of different kinds of smile. *Speech Communication, 50*(4), 278-287.

(3) Ohala, J. J. 1980. The acoustic origin of the smile. *J. Acoust. Soc. Am.* 68.S33.

(4) Papeo, L., Corradi-Dell'Acqua, C., & Rumiati, R. I. (2011). "She" is not like "I": The tie between language and action is in our imagination. *Journal of Cognitive Neuroscience, 23*(12), 3939-3948.

(5) Quené, H., Semin, G. R., & Foroni, F. (2012). Audible smiles and frowns affect speech comprehension. *Speech Communication.*

(6) Schroder, M. (2006). Expressing degree of activation in synthetic speech. *Ieee Transactions on Audio Speech and Language Processing, 14*(4), 1128-1136.

(7) Tartter, V. C., & Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America, 96*(4), 2101-2107.

Sorry, I just don't get it ...

Rosemary Orr

September 9, 2012

1 Introduction

While laughter can be elicited by various means, including laughing gas, (threatening to) tickle, stand-up comedy and joke-telling, it most commonly occurs in verbal exchanges between people in their daily lives. Not everyone can be funny on demand, but pretty much everyone laughs and generates laughter in general social interaction.

2 Why do we laugh with each other?

Broadly speaking, verbal exchanges that contain laughter are generally those in which all participants consider themselves to be in a cooperative social situation where they share common ground.

In departmental or committee meetings, or classroom discussions, we assume that our fellow participants are working towards a common goal: election to the chair, the allotment of funding to research students, transfer of knowledge. Laughter in these situations underlines the common goal, and the shared intention to achieve it. It also releases tension and acts as a distractor when goals, or strategies for achieving them, threaten to conflict.

Laughter in personal relationships is – as well as a sign of cooperation and shared goals – also a reference to the shared experience and shared opinions which reinforce and strengthen these relationships. It signals intimacy, solidarity, affection.

Whether at the workplace or at home, these moments of conversational laughter give a clear message: we understand each other. *We know what we mean.* [1]

In fact, in verbal interactions in new situations with new people, introducing laughter is a message that says: I'm inviting you to laugh with me so that our interaction will be friendly and our relationship productive and enjoyable.

3 Where does it go wrong?

3.1 The academic subculture

At least, that's what I thought until I entered the intercultural subculture of academia about twenty years ago.

The Dublin Laughter Workshop is a multidisciplinary workshop, inviting work on laughter from the points of view of various disciplines, including phonetics, linguistics, psychology, conversation analysis, and human-machine interaction. I notice, with a raised eyebrow, that intercultural marriage counselling and conflict mediation in the international research lab are not on the list.

[1] Of course, this kind of laughter is particularly bonding when we know that people outside the conversation would not understand why we are laughing in the first place ... but there are words for that kind of behaviour ...

When a person enters a relationship, whether professional or personal, in which they are of a different cultural background to the other(s), laughter might not be the silver tinkle of bonding or solidarity, whether in the bedroom, the barroom or the boardroom

While I had always unconsciously subscribed to the notion that the genial aspect of laughing derives from "the binding of companions laughing together in the mutual realisation of safety" [2] , perhaps I should have acquainted myself better with van Hooff's observation that, while a smile might be a sign of appeasement, laughter, with its open mouth and bared teeth, signals dominance and might even stem from "the savage shout of triumph and the cruel mockery over a conquered enemy".

Perhaps more attention to perceptions of the appropriateness of laughter might also be prudent. The academic culture is a serious business where serious people do serious work to attain serious goals. I refer to the weighty tasks of generation and testing of new knowledge, the driving forward of the wheel of progress, the investigation and propagation of civilisation. Frivolity is not the order of the day. As Somerset Maugham pointed out: "Make him laugh and he will think you a trivial fellow, but bore him in the right way and your reputation is assured" [1]

3.2 Wheels within wheels

Any research group worth its citations will owe much of its creativity to influences from beyond the pale. But what are we to make of the cultures within this subculture?

A brief search on the internet will quickly confirm that Germans and women have no sense of humour. If, for example, Japanese or Vietnamese women *do* have a sense of humour, must they cover their mouth to make sure nobody notices? Or is it just inappropriate for women to open their mouths at all?

The particularly Irish custom of *slagging* is not on the list of ten most effective ways to make your Dutch colleagues feel included and valued, and yet *not* being slagged in Ireland – often misconstrued as a sign of politeness – is a very clear message that says "not in our gang".

4 What to do, then?

In my talk, I will outline some of the cultural approaches to laughter and humour and suggest some ways in which we try to circumvent the communications that might arise in intercultural relationships in the academic world. I shall endeavour not to make sweeping statements and generalisations, and to avoid anecdotal evidence. However, since that is the stuff upon which this issue is built, with which it is reinforced, and without which it cannot persist ... I make no promises.

References

[1] W. Somerset Maugham. *The Gentleman in the Parlour: A Record of a Journey from Rangoon to Haiphong*. Random House, 2010.

[2] J. A. R. A. M. van Hooff. A comparative approach to the phylogeny of laughter and smiling. In R. A. Hinde, editor, *Non-verbal Communication*, page 211. Cambridge University Press, 1975.

Dublin, October 26-27, 2012

Third Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisations in Speech

Temporal and spatial patterns in face-to-face conversation

Jens Edlund (KTH Stockholm)

From a conversational point of view, laughter differs from speech in that whereas speech is predominantly one-speaker-at-a-time (although overlaps are certainly not uncommon), laughter is often produced simultaneously – it is indeed said to be contagious. With this as a starting point, I'll present an overview of how various phenomena in spoken face-to-face conversation are temporally and spatially related, and attempt to draw some parallels to laughter.

Temporal distribution of laughter in conversation

Francesca Bonin, Nick Campbell, Carl Vogel
Trinity College Dublin, Ireland

1 Introduction

Laughter, as component of social interaction, has attracted interest within conversational analysis [2, 5]. While laughter can be expressed in different contexts, voluntary or involuntary [6], and diverse in function and degree of functionality [2], it is not random. We study timing of laughter during conversation in relation to topic changes: whether recurrent patterns in laughter distribution with respect to topic changes exist; whether laughter is a reliable topic termination cue. Others also approach this problem [3, 4]. Regularities have been analyzed in the occurrence of shared and not-shared laughter and their different conversational functions [3]. From a large collection of instances two persistent patterns are noted: shared laughter is often associated with topic termination and solo laughter, topic continuation. It has been observed that laughter invites reciprocal laughter [5]; however, Holt qualifies this with analysis of cases in which the listener seemingly refuses the laugh-invitation by continuing the topic with further information, instead.

Keeping in mind Holt's analysis [3, 4], we explore a corpus of multiparty spontaneous chat[1] approaching the problem in two steps: at a coarse-grained level, we analyze the temporal distribution of laughter with respect to topic boundaries; then, at a finer level we will analyze the differences in distribution of shared and solo laughter. The two main points of our work can be summarized by these two questions: I) how laughter is distributed around topic boundaries? II) is there evidence of the "shared laughter-topic termination" relation and of "solo laughter-topic continuation" relation?

2 Distribution of Laughter: analysis

In order to answer to (I), we analyze the temporal distribution of laughter in the before mentioned corpus. The total number of laughter is 713, counting shared and solo laughter. We examine the left and right sides of topic boundaries. Holding the topic change as the central event (hereafter, T-event), we individuate the position of the last laugh (LL) as the last laugh in the previous topic preceding T-event, and the first laugh (FL), as the first laugh following T-event in the new topic (Fig. 1 - left). Given this structure, we calculated the temporal distances (μ) between LL and T-event (μ(LT)) and FL and T-event (μ(TL)), noticing that LLs tend to occur at a shorter temporal distance from the T-event, than FLs. In other words, LT segments are statistically significantly shorter than TL segments (Fig. 1 - right).[2] In this corpus, laughter is more likely as the temporal (and content) distance from the topic boundary increases.

Addressing (II), recall that Holt 2010 [3] notes a clear distinction between shared laughter and solo laughter[3]. Shared laughter is linked with topic termination: it cannot be considered as an independent topic-closing cue, but it may be a supplemental indicator of a topic closing when it occurs in a sequence that is already potentially termination relevant. We repeated the previous analysis of μ(LT) vs μ(TL), distinguishing shared (SH) vs solo (SO) laugh. We focus on the topic termination left neighborhood (μ(LT)). Results of this analysis are reported in Table 2: the median distance between SH laughter and T-event is 4 sec, against the 13 seconds median distance between SO and T-event. SH laughter, rather than SO, tends to occur near a topic termination, and seems to fall in the time-frame that represents Schegloff's termination exchange sequence [7]. Thus, we can argue (supporting Holt) that given a topic termination, it is more likely to find a SH rather than a SO laughter in the termination exchange sequence. Again, this does not mean that SH are sufficient to cue topic termination, but their presence can be a further indicator of a topic termination sequence.

[1] The corpus [1] records conversation in English, including non-natives, among five individuals over three sessions. To our purpose all three days have been used for a total length of about 3h 30, 31523 tokens and 5980 turns. Transcripts present a specific tag for laugh (@w)

[2] One tail wilcox.test, alternative less: p-value = 2.418e-11.

[3] In this study, we define shared laughter situations in which at least two speakers overlap laughing.

Class	Mean	Median
SH-LT	6.36 sec	4 sec
SO-LT	27.58	13 sec
SO-LT U SH-LT	12.75 sec	7 sec

Table 1: LT distances wrt SH and SO laugh

Class	Mean	Median	Position
SH	1.32	0	WI
SO	0.9	0.5	WI
SH	2.7	2	WO
SO	1.19	1	WO

Table 2: SH vs SO distribution in wi and wo

Third Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisations in Speech

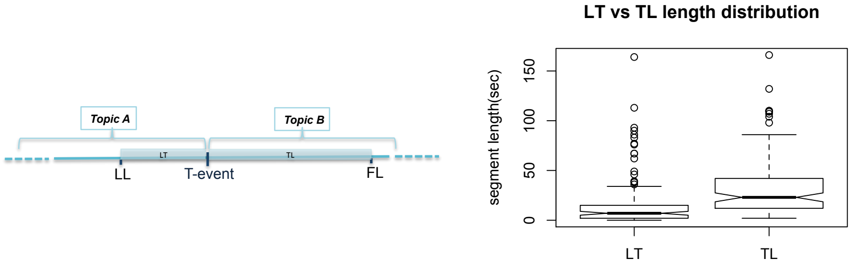

Figure 1: Topic boundary neighbourhood (left) and LT-TL comparison (right)

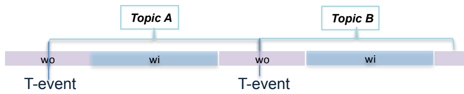

Figure 2: Inter/intra topic segmentation

The second statement in Holts analysis is the relation between solo laughter and topic continuation. In order to investigate this, we analyze the distribution of solo laughter, exploring whether it is more likely to find a SO rather than a SH in relation with a topic continuation segment of the conversation. We divide the corpus in intra topic sections (*wi* segments) and inter topic sections (*wo* segments), where *wi* are defined as the central half of a topic (by definition, those segments do not include a topic change), and *wo* sections as the remaining segments of the corpus overlapping a topic boundary (Fig. 2). If solo laughter are related to topic continuation, we should expect an higher number of SO in topic continuation segments (*wi*), rather than in topic transition segments (*wo*); however, this is not the case (2): there is no significant difference in the distribution of SO laughter among *wi* and *wo* sections. Moreover there is no significant difference between the distribution of SH laughter and SO laughter in intra topic segments, meaning that both (SH and SO) can equally occur in the context of a topic continuation.

3 Conclusions

With respect to I), we find an higher probability of finding a laughter as the distance from the topic boundary increases. With respect to II), we notice that shared laughter tends to occur as topic terminations approach, more than solo laughter; although neither shared nor solo laughter are reliable indicators of topic termination in isolation, shared laughter, more than solo, can contribute (with other features) to constitute a topic termination exchange. Finally, we did not find clear evidence supporting a relation between solo laugh and topic continuation; on the contrary, shared laughter seems to be equally followed by topic continuation utterances. Further studies could be conducted for exploring the nature of those solo laughter (invitation to laugh [5], embarrassment [2]). Next steps will involve also the analyses of different corpora.

References

[1] Nick Campbell. An audio-visual approach to measuring discourse synchrony in m ultimodal conversation data. In *Proceedings of Interspeech 2009*, 2009.

[2] P. Glenn. *Laughter in Interaction*. Studies in Interactional Sociolinguistics. Cambridge University Press, 2003.

[3] Elizabeth Holt. The last laugh: Shared laughter and topic termination. *Journal of Pragmatics*, 42(6):1513–1525, June 2010.

[4] Elizabeth Holt and Paul Drew. Figurative pivots: The use of figurative expressions in pivotal topic transitions. *Research on Language and Social Interaction*, 38(1):35–61, January 2005. UoA 57 (English Language and Literature).

[5] Gail Jefferson. A technique for inviting laughter and its subsequent acceptance/declination. In G Psathas, editor, *Everyday language: Studies in ethnomethodology.*, pages 79–96. Irvington Publishers: New York,NY, 1979.

[6] Mario F. Mendez Mendez, Tomoko V. Nakawatase, and Charles V. Brown. Involuntary laughter and inappropriate hilarity. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 11:253–258, 1999.

[7] E.A. Schegloff. *Sequence Organization in Interaction: Volume 1: A Primer in Conversation Analysis*. Sequence Organization in Interaction: A Primer in Conversation Analysis. Cambridge University Press, 2007.

Laughter in conversational speech: laughing together vs. laughing alone

Khiet P. Truong (1) and Jürgen Trouvain (2)
(1) Unviersity of Twente, The Netherlands
k.p.truong [at] utwente.nl
(2) Saarland University, Saarbrücken, Germany
trouvain [at] coli.uni-saarland.de

Besides spoken words conversational speech usually contains non-verbal vocalisations such as laughter and coughing. In a recent analysis of several publicly available conversational speech corpora (both multiparty and dyadic conversations) we could show that laughter and (other) breathing noises were the most frequent non-verbal vocalisations [1]. What makes laughter even more special, in addition to the frequency in conversations, is the fact that interlocutors often apply laughter as a joint vocal action which is in contrast to most other vocalisations.

Most remarkably, laughter that appears as an utterance of one single speaker ('solo laughter') often shows a different acoustic make-up to laughter where people laugh together. These temporally (partially) overlapping laughs are stronger prosodically marked than non-overlapping ones, in terms of higher values for duration, mean F0, mean and maximum intensity, and the amount of voicing. This effect is intensified by the number of people joining in the laughter event, which suggests that entrainment is at work. We also found that group size affects the amount of overlapping laughs which illustrates the contagious nature of laughter. Finally, people appear to join laughter simultaneously at a delay of approximately 500 ms: this means that spoken dialogue systems have some time to decide how to respond to a user’s laugh.

[1] Trouvain, J. & Truong, K. 2012. Comparing non-verbal vocalisations in conversational speech corpora. Proc. 4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals, Istanbul, pp. 36-39.

On the acoustic vicinity of (adult) crying and (song-like) laughing

Jürgen Trouvain
Saarland University, Saarbrücken, Germany
trouvain [at] coli.uni-saarland.de

Belin et al. [3] noticed in their database of emotional vocalisations an acoustic similarity between the samples portraying the categories "sad" (realised as crying) and "happy" (realised as laughing). This observation is in line with anecdotal evidence of many people who felt unsure whether somebody was crying or laughing when visual and other context information was missing. This is a situation which is highly irritating given the fact that crying is usually associated with negative feelings, and laughing often with positive emotions. This study has two aims: i) a comparison of selected acoustic parameters in the samples of crying and laughing in the above mentioned database [3], ii) the manipulation of cries in order to elicitate the impression of laughter.

To our knowledge there is a lack of phonetic comparisons between (adult) crying and laughing vocalisations, be it on the level of acoustics, perception or vocal production. A notable exception is the study of Erickson [5] investigating laughing, smiled and sad *speech* (not vocalisations). They too, found that "sad and happy speech were very similar in terms of acoustics and articulation" and that their "perception results showed confusion between smile and sad speech." This research gap seems to be true for adult crying in general. Leading experts on crying research such as Vingerhoets states that he is not "aware of any studies analysing the acoustical features of adult crying and the extent to which these features are similar to, or different from, child and infant crying." [11] Although there is a number of studies focusing on the acoustic characteristics of infant crying experimental studies controlling selected acoustic parameters such as F0 range are exceptions (e.g. [7]).

The data for this study were taken from the Montreal Affective Voice Database [3] with 10 actors portraying the emotional categories happiness, sadness, fear, anger, pleasure, pain, surprise, and disgust as vocalisations. Listeners selected the best examples for each speaker as prototypes. Our analysis concentrates on the 10 vocalisations (one per actor) for happy and sad, respectively (resulting in 20 vocalisations). The acoustic analysis reveals a rhythmical similarity between the expressions of both categories: in 19 of 20 vocalisations we find staccato-like quasi-"syllabic" structure [9, 10] with "happy" portrayed as song-like laughter and "sad" expressed as crying. For tempo – quantified here as "call"-rate, a "call" as equivalent to an articulated syllable [2, 9] – we see that laughter (mean 5.3 calls/sec) for all subjects is faster than crying (mean 3.4 syll/sec). In addition, laughter utterances are shorter than crying utterances for 8 out of 10 subjects (mean 1.446 sec vs. 2.229 sec). The mean fundamental frequency is higher for laughter and crying compared to "neutral". Sometimes values up to 500 Hz for male voices and up to 700 Hz for female voices are found. Although there is a tendency for crying showing a higher pitch than laughter, there are great inter-individual differences. A remarkable detail is that laughter often shows one intensity peak whereas crying often reveals two intensity peaks. Regarding mean intensity values and intensity contours no clear differences were found, however there is frequently a declination pattern for intensity and F0.

In the second part of the study the "sad" vocalisations of the database were manipulated to make them confusable with "happy" vocalisations. Informal listening tests with free answers showed that signal manipulations were successful where local tempo adapations were applied by i) shortening the duration of the consonants and vowels according to the mean duration of their laughing counterparts and ii) keep only the more intense one of both intensity peaks. A global adapation in a linear way was not successful.

Although these preliminary results are promising for new insights of crying and laughing as acoustic phenomena it remains largely unclear in which details both vocalisation categories differ in order to maintain the huge contrast in valence as one of the important emotional dimensions. Although the presented ideas could be useful for generating laughter and crying for expressive speech synthesis [4] detailed knowledge about the effects for listeners (or users) is rather limited, e.g. the complexity of laughter as a "happy"-vocalisation [8, 9, 10] or the impression of authenticity [6]. However, laughing and crying should not be reduced to affective reflexes. There is evidence that laughter and crying are optimal carrier for memorising spoken information [1] thus bearing a potential beyond non-verbal communication.

One need for future research on adult crying and laughter is diversity of data and its elicitation. As Vingerhoets et al. [11] put it: "Crying is a rather rare behaviour, that is not easily induced in ethically acceptable ways. Work on crying would be enriched by naturalistic observations of crying behaviour." The main problem here remains the high degree of reluctance (of adults) to cry in public. Consequently it can be hard (though not impossible) to find authentic adult crying in corpora of speech or natural data from mass media. In the workshop we present examples from public TV with negative connotation (loss of the partner) as well as positive connotation (Olympic gold medal winner).

References

[1] Armony, J.L., Chochol, C., Fecteau, S. & Belin, P. 2007. Laugh (or cry) and you will be remembered: Influence of emotional expression on memory for vocalizations. *Psychological Science* 18 (12), pp. 1027-1029.

[2] Bachorowski, J.-A., Smoski, M. & Owren, M. J. 2001. The acoustic features of human laughter. *Journal of the Acoustical Society of America* 111 (3), pp. 1582-1597.

[3] Belin, P. Fillion-Bilodeau S. & Gosselin F. 2008. The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods* 40 (2), pp. 531-539.

[4] Campbell, N. 2006. Conversational speech synthesis and the need for some laughter. *IEEE Transact on Audio, Speech, Language Processing* 14 (4), pp. 1171-1178.

[5] Erickson, D., Menezes, C. & Sakakibara, K. 2009. Are you laughing, smiling or crying? *Proc. Annual Summit and Conference of the Asia-Pacific Signal and Information Processing Assoc.*, pp. 529-537.

[6] Kipper, S. & Todt, D. 2003. The role of rhythm and pitch in the evaluation of human laughter. *Journal of Nonverbal Behaviour* 27, 255-272.

[7] Protopapas, A. & Eimas, P.D. 1997. Perceptual differences in infant cries revealed by modifications of acoustic features. *Journal of the Acoustical Society of America* 102 (6), pp. 3723-3734.

[8] Sauter, D. 2010. More than happy: The need for disentangling positive emotions. *Current Directions in Psychological Science* 19, 36-40.

[9] Szameitat, D.P., Alter K., Szameitat A.J., Wildgruber D., Sterr A. & Darwin, C.J. 2009. Acoustic profiles of distinct emotional expressions in laughter. *Journal of the Acoustical Society of America* 126 (1), pp. 354-366.

[10] Trouvain, J. 2003. Segmenting phonetic units in laughter. *Proc. 15th International Congress of Phonetic Sciences (ICPhS)*, Barcelona, pp. 2793-2796.

[11] Vingerhoets, A.J.J.M., Bylsma, L. & Rottenberg, J. (2009). Crying: A biopsychosocial phenomenon. In: T. Fogen (Ed.), *Tears in the Graeco-Roman world*. Berlin & New York: de Guyter, pp. 439-475.

Dr. Harry Witchel
Brighton and Sussex Medical School
Medical Research Building
Falmer
Brighton BN1 9PS
h.witchel@bsms.ac.uk

Laughing and coughing: testing for vocalised indicators of entrainment and action inhibition

Identifying objective indicators of engagement and disengagement is an important goal within social signal processing and human-computer interactions. One theoretical indicator of engagement in a listener/addressee is action inhibition. One potential action that is inhibited as an intrusion may be coughing. Here we outline a series of “in the wild” experiments structured to test whether the suppression of coughing can be associated with the quality of a live or recorded lecture.

During engagement of dyads the opposite of action inhibition can also occur, when the addressee joins in. However, adding to a conversation may be a sign of politeness or of disagreement. Laughter is generally viewed as a positive signal, although some laughter can be polite. Here we outline an “in the wild” experiment structured to test whether laughter of men or women can be an indicator of attraction during a speed dating session, allowing for us to test the theory that women are attracted to men who make them laugh, and men are attracted to women who laugh at their jokes..

Listen to my breath: how does it sound like? Breathe with me: does it improve emotional attunement?

Raffaella Pellegrini, PhD & Maria Rita Ciceri, PhD
Catholic University of Sacred Heart, Milan, Italy
(raffaella.pellegrini@unicatt.it)

Introduction: Several psycho-physiological studies provided evidences on the influence of psychological variables (such as cognitive processes, performance management and emotional experience) on respiration. Anyway, previous investigations relied most of all on physiological measurements. We argue that also the investigation of the expressive role of breath sounds could be relevant from a psychological point of view since they could have a role in emotional expression and emotional attunement. Previous studies that have addressed the relation between emotions and respirations provided evidences for quite distinct respiratory patterns associated to specific basic emotions (Boiten et al, 1994; Boiten, 1998; Philippot et al, 2002) and, what's more, that mimicking such patterns induce correspondent emotional feeling state (Philippot et al, 2002). This technique is used in some counselling and therapeutic context to reinforce rapport (Bandler et Grinder, 1975; Sutton, 2002; Siegel, 1984) but few investigations have addressed this issue. If "breathing together" could truly enhance emotional responding, that could provide significant cues to be used either in therapeutic setting, interpersonal relations and also in dealing with persons with highly compromised communicative skills (Plotnik et al, 2010).

Aims: The present study aims to investigate whether it is possible to indetifiy distinctive acoustic breathing patterns related to different emotional conditions (anger, fear, sadness, disgust, joy and tenderness) and to investigate how "breathing together" influences the attunement process between participants, considering different dimensions: emotional decoding, similarity of the emotional experiences, perspective taking and interpersonal synchrony. In particular we hypothesize that the more synchronized the imitation of the partner's breathing, the more accurate the understanding of his emotional experience.

Procedure 40 women randomly coupled in 20 pairs voluntarily took part to the study. Six narratives, pre-tested for emotional valence and intensity (anger, sadness, fear, disgust, tenderness, joy), were used as a mean of emotional inductions. First, a 90 sec baseline of participants' breathing at rest was audio recorded. Then participants were asked to read alternatively the narratives and to put themselves into the character shoes. The reader (*identification role*) was asked to breath as if she actually were in that situation while her partner (*mirroring role*) had to listen to her breathing and to express her closeness breathing together with her, in the same way. Both participants kept their eyes closed. Separate tracks of participants' breath sounds were audio-recorded for 90 sec. After each task, they both filled in a questionnaire. Two different versions were written for the mirroring and the identification role that investigated the emotions felt while performing the task, participants' perspective taking and emotional decoding ability. Then, they exchange their role and read the next story, following this procedure for 6 times.

Analyses: *1. Acoustic analysis of breathing tracks:* Audio recording of breathing sounds provided information about features of distinct emotional breathing patterns and about interpersonal synchrony between participants in the attunement task. 420min of breathing audio-tracks were collected. All

RESPs sample of acoustic tracks (210 min) undergo a multilayer analysis. A compound set of measurements that enable a reliable respiratory and acoustic description of breath sounds, as well as to relate partner's respiratory behaviour during the joint task was used (Pellegrini & Ciceri, 2012). *Respiratory indices* includes conventional measurements of temporal features of the respiratory signal; *Acoustic indices*, describe breathing sounds features in particular intensity and timbre features; finally, *Coordination Indices* calculate the lag between couple of participants closest breaths and the number of breaths that fall within 4 progressive thresholds of synchrony. *2. Emotions identification accuracy:* both raw and unbiased hit rates were extracted for each subject as a measure of identification accuracy. 3. Self reports ratings, respiratory, acoustic and coordination indices underwent descriptive analysis, analysis of variance & contrasts analyses.

Results: The study yielded two relevant findings: first it was possible to draw detailed acoustic descriptions of breathing patterns related to distinct emotions. In particular three groups with similar features emerged: 1. Anger & Fear, 2. Tenderness & Baseline; 3. Joy, Disgust & Sadness. Secondly, breathing together influenced many of the attunement dimensions under investigation: 1. Participants closely matched up the timing of their partner's breathing and became more able as time goes on; in particular, interpersonal synchrony seemed to be related to emotional responding and sense of interpersonal similarity but not to emotion identification accuracy. 2. Participants were more able to identify emotional valence than specific emotions: negative ones (fear in particular) were better identified than positive ones. 3. Participants were able to predict their partner experience and they tended to feel a sense of interpersonal similarity and to experience similar emotions.
Concluding, we believe that this research field could provide new, significant understanding to the field of both affective and communication psychology and it could produce effective knowledge and applications to be used in therapeutic settings and interpersonal relations management.

References

Bandler, R. & Grinder, J. (1975b). *The Structure of Magic II: A Book About Communication and Change.* Palo Alto, CA: Science & Behavior Books.

Boiten, F.A (1998). The effects of emotional behaviour on components of the respiratory cycle. *Biological Psychology 49,* 29 – 51

Boiten, F.A., Frijda, N.H. & Wientjes, C.J.E., (1994). Emotions and respiratory patterns: review and critical analysis. *Internationl Journal of Psychophysiology, 17,* 103 – 128.

Philippot, P., Chapelle & G., Blairy, S. (2002). Respiratory feedback in the generation of emotion. *Cognition and Emotion, 16(5),* 605-627

Pellegrini, R. & Ciceri, M.R. (2012). Listening to and mimicking respiration: Understanding and synchronizing joint actions. *Review of Psychology*, *in publication.*

Plotnik, A., Sela, L., Weissbrod, A., Kahana, R., Haviv, L., Yeshurun, Y., Soroker, N. & Sobel, N. (2010). Sniffing enables communication and environmental control for the severely disabled. *Proceeding of the National Academy of Sciences, 107(32),* 14413-14418.

Siegel, E. (1984*). Dance Movement Therapy: Mirror of Ourselves: The Psychoanalytic Approach.* New York: Human Science Press.

Sutton, J. (2002). Preparing a potential space for a group of children with special needs. In A. Davies & E. Richards (Eds.), *Music Therapy and Group Work: SoundCompany.* London: Jessica Kingsley.

Index of Authors

Notes