Proceedings of Laughter Workshop 2018

Jonathan Ginzburg, Catherine Pelachaud (eds.)

Sorbonne Université, September 2018

SPONSORS

LLF
Laboratoire de linguistique formelle

Labex EFL

SORBONNE PARIS CITÉ

IUF
institut
universitaire
de France

SORBONNE
UNIVERSITÉ

cnrs

ISIR
INSTITUT
DES SYSTÈMES
INTELLIGENTS
ET DE ROBOTIQUE

We are happy to present Laughter Workshop 2018, the 5th triannual workshop on Laughter. This year's workshop is hosted at Sorbonne Université, named for the great theologian and founder of the Parisian university, but not apparently a man with a great laugh. Laughter Workshop 2018 continues the tradition of presenting high-quality talks and posters on laughter from a variety of perspectives such as phonetics, cognitive neuroscience, formal semantics and pragmatics, cognitive psychology, and artificial intelligence.

17 submissions were received for the main session, and each was reviewed by two experts. 10 talks were selected for oral presentation; the poster session hosts most of the remaining submissions.

We are lucky to have three world famous researchers as invited speakers—Sophie Scott and Gary McKeown. They both represent a broad range of perspectives and disciplines. We are sure that their talks will stimulate much interest and (hopefully) at least some controversy. Together with the accepted talks and posters we look forward to a productive and interactive conference.

We are grateful to the reviewers, who invested a lot of time giving very useful feedback, both to the program committee and to the authors, and to members of the local organizing committee, Reshma Kantharaju, Eimear Maguire, and Chiara Mazzocconi for their hard work in helping to bring the conference to fruition.

We are also very grateful to a number of organizations, who provided generous financial support to Laughter Workshop 2018:

- Institut Universitaire de France
- ISIR, Sorbonne Université
- Laboratoire de Linguistique Formelle, Université Paris-Diderot
- The Laboratoire d'excellence LabEx-EFL (Empirical Foundations of Linguistics), Paris Sorbonne-Cité.

Jonathan Ginzburg, Catherine Pelachaud
September, 2018

Programme Committee

- Nick Campbell (School of Linguistic, Speech and Communication Sciences, Trinity College, Dublin)
- Dirk Heylen (Human Media Interaction, University of Twente)
- Jonathan Ginzburg (LLF, Université Paris-Diderot)
- Catherine Pelachaud (CNRS ISIR, Sorbonne Université)
- Khiet Truong (Human Media Interaction, University of Twente)
- Jürgen Trouvain (Computational Linguistics and Phonetics, Saarland University)

Local Organizing Committee

- Jonathan Ginzburg (LLF, Université Paris-Diderot)
- Reshmashree B Kantharaju (ISIR, Sorbonne University)
- Eimear Maguire (LLF, Université Paris-Diderot, Co-chair)
- Chiara Mazzocconi (LLF, Université Paris-Diderot, Co-chair)
- Catherine Pelachaud (CNRS ISIR, Sorbonne Université, Co-chair)

CONTENTS

Voluntary and Involuntary Mechanisms in Laughter Production and Perception

Sophie Scott

University College, London

In this talk I will compare and contrast the vocalisation mechanisms that (hypothetically) underlie different kinds of laughter production. I will extend this think about a continuum between spontaneous and more communicative laughter, and address some recent findings on developmental conditions where affected individuals find that these distinctions can be harder to make.

The actions of peripheral linguistic objects: clicks

Richard Ogden
Department of Language & Linguistic Science
Centre for Advanced Studies in Language & Communication
University of York, YORK YO10 5DD, England
richard.ogden@york.ac.uk

Abstract

This paper is a conversation analytic study of the linguistic, phonetic, sequential and multimodal resources participants in conversation have to make sense of clicks in spoken English.

1 Introduction

Non-verbal vocalisations in spoken interaction are often assumed to play an important role in displaying affective stances. This paper will focus on clicks (‘tut tut’ or ‘tsk’ sounds), a vocal but not verbal practice common in English and many other European languages. Clicks have been studied from a conversation analytic perspective, but much is still unknown about the affective work they do, their visual characteristics, and how participants in interaction themselves interpret their contribution to an ongoing conversation. This paper takes a conversation analytic approach to the analysis of clicks in naturally-occurring interactions, and shows what semiotic resources are available to participants to make sense of clicks in one another’s talk.

Clicks make an interesting case for non-verbal vocalisations. Unlike particles like ‘wow’ or ‘aw’, they are not amenable to prosodic manipulation such as duration, or F0 adjustments. Some of them arise from preparations for speaking, and have an iconic interpretation: ‘I am about to speak’ (Ogden 2013). Others, such as those which are the topic of this paper, have a more complex semiosis, and exhibit more linguistic properties.

An important task for participants in conversation is to establish what *action* a co-interactant has implemented in a prior turn. This is known as action ascription (Levinson 2013). In Example 1, D identifies a problem in his arrowed turn ‘I don’t know…’. M at her arrowed turn displays her understanding of this as a request, which she declines. Thus M has *ascribed* to D’s turn the *action* of requesting.

Ex.1 MDE stalled

```
D:   ˙hh My ca:r is sta::lled.
     (0.2)
D:   ('n) I'm up here in the Glen?
M:   Oh::.
     (0.4)
D:   ˙hhh A:nd.hh
     (0.2)
D:→  I don' know if it's po:ssible, but˙hhh see I
     haveta open up the ba:nk.hh
     (0.3)
D:   a:t uh: (·) in Brentwood?hh=
M:→   =Yeah:- en I know you want- (·) en I whoa-
     (·) en I would,
```

The wider research question is: what is the relation between linguistic design of turns at talk and the actions participants may ascribe to those turns? and how should they respond? More specifically for this paper: how do participants interpret clicks, a family of sounds whose linguistic status is marginal, whose semantic content is vague, and whose phonetic form is not amenable to prosodic manipulation? Our focus is on how actions are recognised, rather than which actions are implemented, which is the subject of Ogden (2013).

2 Data

The data for this paper is a collection of 168 clicks extracted from the CallHome corpus. The data are presented in summary in Fig. 1. This data is supplemented with material from other data sets.

The coding combines phonetic and conversation analytic categories, including:

- **Phonetic** features: central vs. lateral airflow; oral vs. nasal airflow; single vs. multiple productions
- **Location in the turn**: standalone, pre- or post-positioned, or mid-turn (Schegloff, 1996)
- **Action**: indexing a new sequence, displaying an affective stance, self-repair, etc.

According to native speaker intuition (and dictionary entries), clicks display disapproval or annoyance (Wright, 2007); but as we will see, an interactional analysis provides a more nuanced view of how standalone clicks function. We will focus on multiple and post-positioned clicks, which have complex meanings.

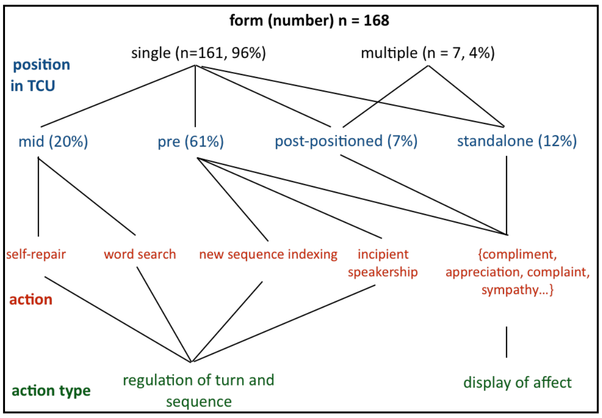


Fig. 1: Distribution of clicks in the data

3 Standalone Clicks

In response to complaints and troubles telling, clicks (!) can occur alongside response particles and/or verbal material in the same turn, as in Ex. 2 below, a complaint about a new manager at a factory.

Ex. 2: CH en_5278.165-186 the factory

```
14 B   also he's Also cOsting a FORtune.
15 A    `Oh gee:. !
16 B    this <<cr> guy.>
17 A    hh° god
```

The verbal material in such turns provides evidence of one of two relevant response types to troubles or complaints: 'displaying sympathy with the teller', or (as here) 'displaying disapproval of the source of the trouble'. Without response tokens or verbal material, the ascription of a particular action in such cases is not trivial; but a standalone click may ambiguously project 'sympathy' or 'disapproval', which are both affiliative and aligning responses. The next two sections illustrate.

3.2 Clicks treated as continuers

One of the commonest sequential environments for standalone clicks is:

1. A speaker produces a turn in which troubles are told or a complaint about a third party is made
2. A recipient produces a click (!) shortly after a Transition Relevance Place in the prior turn
3. The troubles-teller or complainant continues their turn, and in doing so does not treat the click as disruptive, nor as a turn by itself. Rather, the click is treated more like a continuer.

Ex 3-4 illustrate with complaints which are receipted with a click but no verbal material.

Ex. 3: CH en_5254.484-500.dreadful and cold

A is complaining about how her parents in law treated her over Christmas.

```
09 A    =they were really (.) ↓`drEAdful.=
10      =and thE[:n-] and `↑vEry very `cOld.=
11 B            [ ! ]
12 A    =.h [?and you know ?I have just been
13 B        [hm.
14 A    SO devoted and SO loving=
```

Ex. 4: CH en_4822.1078-1093 cancelling

A is complaining about a private student.

```
02 A    [°h] so Anyway i went out and bought
03      all these books and like threw myself
04 A    into it heart and soul and then she
05 A    nEver shows Up.
06 B    ! (-)°h[h ]
06 A           [sh]e's always cAlling and
07      cAncelling or nOt calling and nOt
08      showing an-
```

In such cases, the click does not disrupt the trajectory of the complaint or troubles telling, but is treated by the teller as allowing them to progress with their telling. Another option from the recipient would be a continuer, such as 'uh-huh' or 'mhm', registering continued recipiency without taking an affective stance towards the ongoing talk. This sequence shows that standalone clicks demonstrate an orientation to the *relevance* of a response, and perhaps specifically to an *affect-laden* response, but there is no evidence from the talk itself what kind of affective stance the click delivers.

3.2 Clicks treated as insufficient

Sometimes, a complainant or troubles teller orients to a click as an insufficient response. In these cases, the sequence is a little more complex. The click is immediately followed by an insert from the teller which is an overt request for a display of understanding: 'you know?' or 'you know what I mean?', thus treating the click as too minimal to count as adequate. Interestingly, these cases show that the continuer which follows this request, 'mhm' (lines 10 and 11 respectively), minimal as it is, *is* treated as sufficient for the teller to continue with their telling.

Ex. 5: CH en_5254.932 waitress

```
05 R    .h now if I go back to (Newark)
06      what am I gonna do=be a waitress
07      do [book-keepi[ng
08 L       [!         [{p mm}
09 R    y'know?
10 L    mhm
```

```
11 R    I have NO skills really=
```

In both sequences, clicks are treated as a minimal object. Participants orient to the *minimality* of the response provided by the click, and as it is treated as allowing the talk in progress to continue, it is *aligning* and *affiliative* (Stivers, 2008).

4 Multiple Clicks

Multiple clicks are a deliberate vocalisation. Their rarity makes any conclusive statement about their form or function difficult. Nonetheless, features of their position in a turn, and features of their co-production (such as the time interval between them or accompanying lip rounding) can be recruited meaningfully. The cases here occur post-positioned after a turn, thus serving as a 'post completion stance marker' (Schegloff, 1996, 92-3). In both cases, the rhythmical pulse established by the clicks is recruited by the incoming speaker to time their turn (cf. Ogden & Hawkins, 2015).

4.1 Mirroring

In Example 6, A and B have been discussing a record by Michael Jackson that allegedly contained anti-Semitic lyrics and was withdrawn from sale. B produces multiple clicks in response to A's laughter particles in the service of affiliation with A's stance.

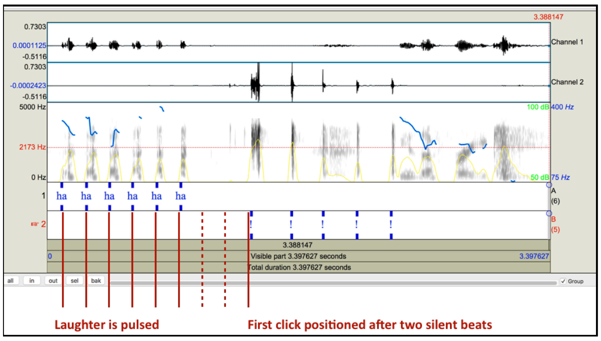
Ex. 6: CH en_4092.1497-1597 michael jackson

```
18 A    would yOU belIEve it,
19     “oh I didn’t know it was of’FENsive?”
20      ha ha ha ha ha ha
21 B    ! ! [ ! ]! !=
22 A        [°h ]
23      =hE’s a ↓`FREAK.((laugh))
24 B    <<p l> yeah he IS.>
```

At lines 18-19, A doubts his claim to innocence, and at l.20 she produces six post-completion laughter particles, taking a mocking stance to his claim. These are followed at l.21 by five clicks from B (Fig. 2), and then a negative assessment of Jackson from A, which B agrees with at l.24. The clicks thus display affiliation with A's stance towards Jackson.

F0 rises through the laughter particles. The clicks have a falling Centre of Gravity (CoG), produced by progressively increasing the lip rounding. The falling 'pitch' of the clicks symmetrically mirrors the rising pitch of the laughter. The laughter pulses are isochronous. The first click of B's response falls on beat (after two silent beats) with the pulse projected by A's laughter particles. The phonetic design of the multiple clicks matches that of the laughter rhythmically and prosodically, despite the fact that clicks are not easily manipulated in the prosodic domain. As Couper-Kuhlen (2012) has suggested, reciprocating the prosody of another is a very basic iconic method for displaying affiliation. While there are plenty of examples of this in verbal material, this example shows that it can also work in non-verbal material, or events which are affiliated with speech.

Fig. 2: Pulsed laughter, on-beat clicks; rising F0, falling CoG



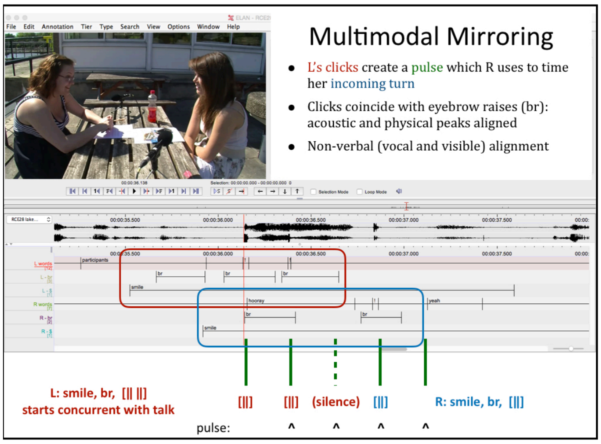
Laughter is pulsed

First click positioned after two silent beats

4.2 Clicks and other modalities

In face-to-face data, clicks are frequently associated with winks, eyebrow flashes, nods or the apex of gestures, i.e. with peaks of physical activity. (Loehr, 2007). Here we consider an example of lateral clicks accompanied by visible behaviours across the turn space.

Fig. 3: Ex. 7. Coordination of clicks, eyebrow flashes (br) and smiles across the turn space.



Multimodal Mirroring

- L's clicks create a pulse which R uses to time her incoming turn
- Clicks coincide with eyebrow raises (br): acoustic and physical peaks aligned
- Non-verbal (vocal and visible) alignment

L: smile, br, [|| ||] starts concurrent with talk

R: smile, br, [||]

L(eft) produces an apparent compliment to R(ight): 'you have the best participants', followed by two lateral clicks [ǁ ǁ] as a post-completion stance marker. While L produces 'participants', she smiles and does an eyebrow flash. These clicks are accompanied by eyebrow flashes. L's smile, the eyebrow flashes and [ǁ ǁ] are reciprocated by R. R's response to L's turn is to reciprocate the lateral click with an eyebrow flash; she thus seems to accept L's comment on her own turn, and to ratify it by mirroring L's own vocal (not verbal) and visible behaviours. Note also that R's click comes in on beat, after a beat of silence, and thereby displays alignment with L.

L's two lateral clicks, along with the other visible behaviours, seem to modify the understanding of 'you have the best participants': they invite R to collude in an understanding that they share but do not verbalise. The implication is that L is one of R's participants, and so her turn is retrospectively self-congratulatory, rather than an 'innocent' compliment. The same affective stance is found with [ǁ ǁ] in other cases, such as (obscene) jokes.

Alongside the clicks, speakers can recruit rhythm, inter-speaker temporal coordination, and facial expression to express something that is not verbalised.

5. Conclusions

I have focused on the ascription of action to standalone and multiple clicks in conversation. Standalone clicks frequently occur in a sequential position where a display of sympathy or disapproval is relevant. The temporal placement of a click soon after a Transition Relevance Place in another speaker's talk displays an orientation to the relevance of a response. Other such displays can involve responses particles and verbal material. They contrast with affectively neutral continuers like 'mhm' in the same position. Standalone clicks, without verbal material in the same Turn Constructional Unit, are ambiguous between displaying sympathy or disapproval, and convey broad affiliation with the complainant or troubles-teller. This minimality makes standalone clicks useful as a resource for displaying affiliation without committing to a particular affective stance.

When post-positioned, clicks are used to adopt an affective stance towards the prior TCU; but the precise interpretation depends on features of the click, such as the whether the click is released centrally or laterally. Multiple clicks provide a metronome-like device for co-participants to coordinate their incoming talk. On-beat talk is commonly an iconic means of displaying alignment and affiliation with another speaker. In addition, other embodied behaviours such as smiles and eyebrow flashes are an important part of the design of the click construction; these co-occurring bodily behaviours provide participants with a multimodal set of semiotic resources.

REFERENCES

Couper-Kuhlen, E. (2012). Exploring affiliation in the reception of conversational complaint stories. In A. Peräkylä & M.-L. Sorjonen (Eds.), *Emotion in Interaction*. (pp. 113–146). Oxford, New York: Oxford University Press.

Levinson, S. C. (2013). Action Formation and Ascription. In J. Sidnell & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 103–130). Chichester: Blackwell.

Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, 7(2), 179–214. http://doi.org/10.1075/gest.7.2.04loe

Ogden, R. (2013). Clicks and percussives in English conversation. *Journal of the International Phonetic Association*, 43(3), 299–320. http://doi.org/10.1017/S0025100313000224

Ogden, R., & Hawkins, S. (2015). Entrainment As a Basis for Co-Ordinated Actions in Speech. In *Proceedings of ICPhS XXVIII*. Glasgow. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html

Schegloff, E. A. (1996). Turn Organization: One Intersection of Grammar and Interaction. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and Grammar* (pp. 52–133). Cambridge: Cambridge University Press.

Stivers, T. (2008). Stance, Alignment, and Affiliation During Storytelling: When Nodding Is a Token of Affiliation. *Research on Language & Social Interaction*, 41(1), 31–57. http://doi.org/10.1080/08351810701691123

Wright, M. (2007). Clicks as markers of new sequences in English conversation. In *International Congress of the Phonetic Sciences XVI* (pp. 1069–1072). Saarbrücken. Retrieved from www.ichps2007.de

Classification and clustering of clicks, breathing and silences within speech pauses.

Anna Canal Garcia, Marine Collery, Velisarios Miloulis,
Zofia Malisz
KTH, Stockholm

Abstract

This work reports on automatic classification of conversational events that occur in pause intervals between speech activity. The available classes are: audible breathing, oral clicks and silences. We implement a supervised algorithm (SVM) on labeled data of one speaker with 92% testing accuracy on the same speaker. Additionally, we explore unsupervised methods such as DBSCAN with t-SNE-based dimensionality reduction on that speaker and a large conversational corpus.

1 Introduction

Pause intervals between speech activity in conversation harbour a variety of phenomena (Trouvain, 2014). These events, for example: breathing, hesitations and oral clicks provide information about the dialogue state and turn-taking (Włodarczak and Heldner, 2016) and about the linguistic processing state or speech planning processes (Wright, 2011).

Discourse clicks (Wright, 2011; Gold et al., 2013; Ogden, 2013; Ward, 2006; Trouvain and Malisz, 2016) are usually perceptually salient, unconsciously produced, isolated sound events occurring within pause intervals. They are articulated by forming an anterior closure e.g. with the tip of the tongue against the alveolar ridge. The release of the closure generates a burst-like sound.

Such clicks can be understood as discourse markers with the function of either indexing a new sequence or signaling lexical access difficulties. They are often co-located with fillers, i.e. hesitation particles in ”filled pauses” such as ’uh’ or ’uhm’ in English. Spontaneous in-pause clicks can also be understood as a speech preparation gesture (Scobbie et al., 2011), particularly given their vicinity to breathing (Wright, 2011; Trouvain, 2014). Given all these factors, clicks, like inhalation noise and fillers, are good indices of the depth of a prosodic break.

In the present paper, we ask whether it is possible to automatically distinguish breathing noises and oral clicks from each other and from ”true” silences in pause segments of spontaneous speech. There are automatic methods to detect breathing that perform with high accuracy (Braunschweiler and Chen, 2013; Fukuda et al., 2011). However, we are only aware of few studies detecting and classifying oral clicks - typically as part of larger sound classification tasks (Temko et al., 2009).

2 Data

The immediate goal of this study is to distinguish non-verbal speech events occurring inside speech pauses, such as breathing, clicking from silence by using supervised and unsupervised methods.

For the event classification and clustering task we use data from a) a case study of a prolifically clicking English speaker and from b) a large conversational corpus in German. The way the datasets are used for the task is sketched in Fig.1.

2.1 Case study

The case study consists of a single, female speaker giving a keynote address (Cutler, 2014) in English (7 audio files and a total of 58m4s, sampling rate = 22050Hz). The speaker produces a large number of conversational clicks in her speech (11 per minute) that are clearly audible. All in-pause events were segmented and manually labeled into 5 in-pause categories: silence, audible inhalation, click events, individual click bursts and fillers by (Trouvain and Malisz, 2016) who also report on adequate inter-annotator agreement in this data.

2.2 Corpus

GECO is a corpus (Schweitzer and Lewandowski, 2013) of 22 German spontaneous dialogues between 11 unique female speakers on topics of their choice (92 audio files, total 1d 15h 56m and 22s recordings, sampling rate = 48kHz). The corpus is delivered with pause segmentation but in-pause events were not identified and categorized. Inspection of the audio showed that the pauses contain some clicks, breathing and silences.

3 Method

We provide an overview of our approach in Fig.1. The code can be found at: `https://github.com/annacanal/Inter-speech-event-classification.git`.

3.1 Acoustic feature extraction

Pauses were extracted directly from the case study data. Each pause segment available in the corpus was split into windows of 25ms before feature extraction. We used openSMILE for acoustic feature extraction. The extracted features and functionals are listed in Table 1. All low level descriptors were processed by a simple moving average (SMA) low-pass filtering. Dynamic features (delta regression coefficients) were added per LLD. The resulting feature set has 338 dimensions. The data was standardized with zero mean and unit variance and all features with zero variance were dropped before applying the algorithms.

Features	**Functionals**
RMS energy, Δ	max, min, range,
MFCC 1-12, Δ	amean, stddev, linregc1,
ZCR, Δ	linregc2, linregerrQ,
(Zero Cross. Rate)	skewness, kurtosis

Table 1: Acoustic features and functionals.

3.2 Dimensionality reduction

We tested the usefulness of TSNE (t-distributed Stochastic Neighbor Embedding). We used TSNE to project the feature space from 338 to 2 dimensions. We tested the performance of different dimensionality of the feature set as input to both supervised and unsupervised tasks. Also, data in 2 dimensions visualizes the distributions well.

3.3 Supervised classification

We used Support Vector Machines (SVM). A meta-analysis in (Pramono et al., 2017) showed that SVM are the most frequently used and successful algorithms in detection and classification of breathing events.

3.4 Unsupervised clustering

We used DBSCAN, a clustering algorithm that showed the best distinctiveness of classes in comparison to other standard clustering methods we tested (KMEANS and AGGLOMERATIVE CLUSTERING). DBSCAN creates an arbitrary number of clusters in areas with high density, and assumes they are separated by areas of lower density. It is governed by two parameters `eps` and `min_samples` which have to be found by a search. The algorithm identifies any points it cannot assign to specific clusters as "noise". Those are reported as "DBSCAN Noise" below.

4 Results

4.1 Supervised classification

4.1.1 Case study

We used three categories from the labeled case study dataset: breaths, click events and silences within the pause segments for training. We trained the SVM with 5-fold cross-validation, the data split used in each fold was: training set (90%) and test set (10%). We measured the accuracy, precision and recall of the inter-speech event classifier in each fold, the mean and standard deviation of these measures are shown in Table 4.1.1. The majority of pauses were correctly classified.

$\mathbf{F_{dim}}$	**Accuracy**		**Precision**		**Recall**	
	M	SD	M	SD	M	SD
338-D	0.92	0.02	0.89	0.04	0.86	0.03
2-D	0.86	0.01	0.80	0.02	0.79	0.04

Table 2: Performance for SVM cross-validation in the case study depending on feature dimension.

To illustrate the correct and the incorrect classifications for each class, we provide confusion matrices on the test set when using 338-D and 2-D in Figure 2. Since we had class imbalance, the matrices had been normalized.

The 338-D feature set improved the classification of silences and breathing relative to the 2-D feature set. However, the classification of clicks is better with data cast to a 2-D feature space. The results also indicate that breathing and clicks are well-distinguishable from each other: in case they are misclassified, they are labeled as silences.

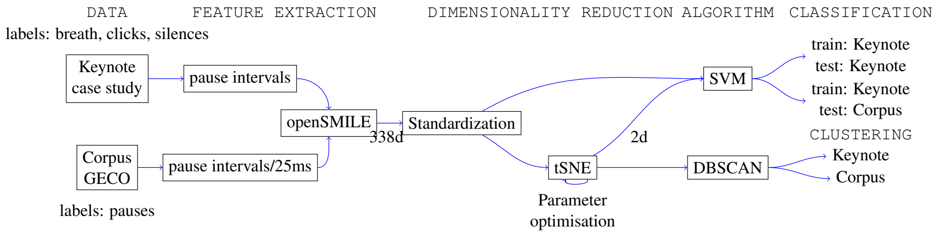


Figure 1: Schematic of the approach

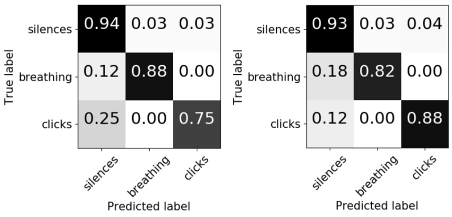


Figure 2: Accuracy confusion matrix for the classes as predicted by SVM in the case study test set. Left: using 338-D, right: using 2-D features.

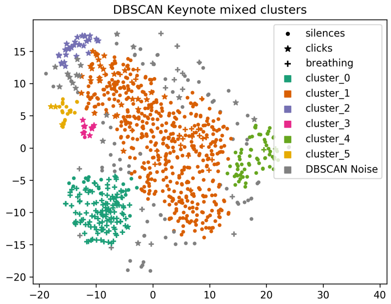


Figure 3: Colors represent clusters predicted by the DBSCAN in keynote data. Different shapes are the actual clusters.

4.1.2 Corpus

We were interested whether the supervised classifier could predict the events of interest in unseen data from different speakers and a different language (English vs. German). We used all the keynote data for training (388-D feature set) and then applied to GECO as a test set. A preliminary manual annotation of the reference classes in GECO was done by an expert in phonetics.

The comparison of the predicted classes with the preliminary gold standard showed that the breathing class was distinguishable from silences in only 50% of cases while the clicks were not up to a classification standard (Acc = 0.16). We obtained low results from the evaluation on all the predictions (Acc = 0.48).

4.2 Unsupervised clustering

4.2.1 Case study

Results of DBSCAN clustering using the case study data are shown in colours in Fig.3. The different point shapes show the actual clusters in this data. The resulting breathing cluster (cluster 0) includes almost all actual breathing events. Some points are included in DBSCAN noise. Cluster 2 encompasses clicks, however, the actual click cluster is larger. Actual clicks are sparse and not easy to cluster; the confusability is expressed in cluster: DBSCAN noise. Finally, the DBSCAN clustering shows a large, homogeneous cluster (cluster 1) that matches the silences cluster as obtained from the full and evaluated gold standard labels for the keynote.

4.2.2 Corpus

Unsupervised results on individual conversations (GECO) showed cases of unspecific clusters of clicks. But it was usually easy to find clusters containing mostly breathing events. One example where clusters matched the data well is conversation A-C (Fig.4, left). Cluster 2 corresponds to clicks and cluster 0 to breathing. The other clusters gather silences, background noises, some breathing, some clicks and unidentified noise.

In conversation D-L (Fig.4, right), cluster 1 and cluster 3 are gathering breathing events, while cluster 0,2 and 4 are mostly silences, unidentified noise and some breathing. No cluster comprising clicks events is distinguishable. The actual low number of clicks events is very likely the reason.

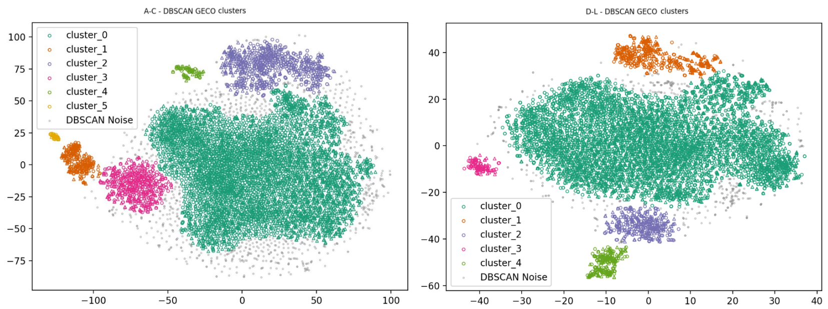


Figure 4: Clusters predicted by DBSCAN in 2 conversations from the corpus data. Left: A-C conversation, right: D-L conversation.

5 Discussion

Our results show that an automatic classification of in-pause events, such as breathing and clicks, is possible. The supervised approach provides solid classification accuracy when trained and tested on the same speaker. It deals with imperfect audio conditions, e.g.: background noise is almost continuous for the keynote. However, the classifier trained on the single speaker of English did not generalise to the German multispeaker test set. In the unsupervised approach, we found that TSNE dimensionality reduction works quite well to differentiate between the events in the feature space.

Even if we did not manage to perfectly separate the clusters, it is apparent that breathing is easier to model than clicks. Clicks are often overclustered and do not form as distinct a cloud as breathing datapoints in most of the studied datasets.

6 Acknowledgements

References

Norbert Braunschweiler and Langzhou Chen. 2013. Automatic detection of inhalation breath pauses for improved pause modelling in HMM-TTS. In *Eighth ISCA Workshop on Speech Synthesis*.

Anne Cutler. 2014. Learning about speech. In *Keynote given at INTERSPEECH 2014*, Singapore.

Takashi Fukuda, Osamu Ichikawa, and Masafumi Nishimura. 2011. Breath-detection-based telephony speech phrasing. In *Proceedings of INTERSPEECH2011*, pages 2625–2628, Florence, Italy.

Erica Gold, Peter French, and Philip Harrison. 2013. Clicking behavior as a possible speaker discriminant in English. *Journal of the International Phonetic Association*, 43(3):339–349.

Richard Ogden. 2013. Clicks and percussives in English conversation. *Journal of the International Phonetic Association*, 43(3):299–320.

Renard Xaviero Adhi Pramono, Stuart Bowyer, and Esther Rodriguez-Villegas. 2017. Automatic adventitious respiratory sound analysis: A systematic review. *PLOS ONE*, 12(5):e0177926.

Antje Schweitzer and Natalie Lewandowski. 2013. Convergence of articulation rate in spontaneous speech. In *INTERSPEECH*, pages 525–529.

James M Scobbie, Sonja Schaeffler, and Ineke Mennen. 2011. Audible aspects of speech preparation. *Proceedings of 17th ICPhS, Hong Kong*, pages 1782–1785.

Andrey Temko, Climent Nadeu, Dušan Macho, Robert Malkin, Christian Zieger, and Maurizio Omologo. 2009. Acoustic event detection and classification. In *Computers in the human interaction loop*, pages 61–73. Springer.

Jürgen Trouvain. 2014. Laughing, breathing clicking–the prosody of nonverbal vocalisations. In *Proc. Speech Prosody*, pages 598–602.

Jürgen Trouvain and Zofia Malisz. 2016. Inter-speech clicks in an interspeech keynote. In *INTERSPEECH 2016*, pages 1397–1401. International Speech Communication Association.

Nigel Ward. 2006. Non-lexical conversational sounds in American English. *Pragmatics & Cognition*, 14(1):129–182.

Marcin Włodarczak and Mattias Heldner. 2016. Respiratory belts and whistles: A preliminary study of breathing acoustics for turn-taking. In *INTERSPEECH 2016, San Francisco, USA, September 8–12, 2016*, pages 510–514. International Speech Communication Association.

Melissa Wright. 2011. On clicks in English talk-in-interaction. *Journal of the International Phonetic Association*, 41(2):207–229.

On Breath Noises – A short Review

Jürgen Trouvain
Dept. of Language Science and Technology, Saarland University,
Saarbrücken, Germany

trouvain@coli.uni-saarland.de

Breath noises as acoustic and audible reflections of inhalation and exhalation are probably the most common non-verbal vocalisations in spoken communication. Breath noises can occur in a multitude of occasions and they can serve as functional markers in various ways.

The prime example are inhalation noises in speech pauses. Here, breath noises function as markers of prosodic-syntactic boundaries, so that the term breath-groups is sometimes used for intonation (or prosodic) phrases (Lieberman 1967). Phonetic studies showed how duration and intensity of the inhalation noises are used for utterance planning in speech production and how they inform listeners about the length of the upcoming phrase (Fuchs et al. 2013). Interestingly, when speakers are under physical stress they show different forms of breath noises in speech pauses, e.g. with extreme exhalation noises (Trouvain & Truong 2015).

Regarding laughter various forms can be described with characteristic noises of exhalation and inhalation (Bachorowski & Owren 2001). A strong inhalation noise can mark the offset of a long and complex laugh (Chafe 2007). Also in (other) affect bursts, breath noises can play a crucial role, such as startle or in crying (Trouvain 2011).

On the level of pragmatics, breath noises can be used as a discourse marker with the intent to take the turn, and in some cultures respiratory noises are markers of politeness, e.g. in Korean (Winter & Grawunder 2012). Breath noises also have a high potential of signaling individuality, either by idiosyncratic acoustics, e.g. by inhalation noises with an ingressive fricative [s] (Trouvain 2010), or by different patterns of inhalation and exhalation (Kienast & Glitza 2003). The (incomplete) list above shows that breath noises are a rather rich source of information on the linguistic but also on the non-linguistic level.

Surprisingly, breath noises are often and maybe systematically ignored in speech analysis, speech synthesis and speech recognition. This is reflected for instance by the fact that in speech fluency research pauses that contain breath noises are regarded as 'silent'. In some conversational corpora the annotation schemes do not have a category for breath noises (Trouvain & Truong 2012). Likewise, speech prosodists regularly ignore breath noises as important acoustic cues of prosodic phrase boundaries.

Pauses in synthesised speech are often not modelled in a human-like way (Trouvain & Möbius 2018) and they virtually never contain breath noises. However, breath noises would be probably beneficial to have speech synthesis in a pleasant and a memorisable way (Whalen et al. 1995) and necessary for expressive speech synthesis. Breath noise in automatic speech recognition is still a low researched topic though there are various approaches for explicit breath detection (e.g. Fukuda et al. 2018).

While there are research groups working on the physiological, particularly the kinematic, fundaments of respiration in speech (e.g. Bailly et al. 2013, Fuchs et al. 2015, Włodarczak & Heldner 2017) the link between kinematic and acoustic signals of inhalation and exhalation in speech is not yet fully understood. Thus, the focus of this review is two-fold: i) on the acoustic characteristics of breath noises found in a variety of speech data: several forms of read speech, speech before and after physical exercise, lectures, radio live commentaries, parliamentary speech and their simultaneous interpretations, and dialogues; ii) on the possible functions of the various forms of breath noises.

References

Bachorowski, J.A. & Owren, M.J. 2001. Not all laughs are alike: voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science* 12(3), pp. 252-257.

Bailly, G., Rochet-Capellan, A. & Vilain, C. 2013. Adaptation of respiratory patterns in collaborative reading. *Proc. Interspeech*, Lyon. pp. 1653-1657.

Chafe, W. 2007. *The Importance of Not Being Earnest*. Amsterdam: Benjamins.

Fuchs, S., Petrone, C., Krivokapic, J. & Hoole, Ph. 2013. Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics* 41, pp. 9-47.

Fuchs, S., Petrone, C., Rochet-Capellan, A., Reichel, U. & Koenig. L. 2015. Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands. *Journal of Phonetics* 52, pp. 35-45.

Fukuda, T., Ichikawa, O. & Nishimura, M. 2018. Detecting breathing sounds in realistic Japanese telephone conversations and its application to automatic speech recognition. *Speech Communication* 98, pp. 95-103.

Kienast, M. & Glitza, F. 2003. Respiratory sounds as an idiosyncratic feature in speaker recognition. *Proc. 15th International Congress of Phonetic Sciences (ICPhS)*, Barcelona, pp. 1607-1610.

Lieberman, Ph. 1967. *Intonation, Perception and Language*. Cambridge, Mass.: MIT Press.

Trouvain, J. 2010. Affektäußerungen in Sprachkorpora. *Proc. 21. Konferenz Elektronische Sprachsignalverarbeitung (ESSV '10)*, Berlin, pp. 64-70.

Trouvain, J. 2011. Zur Wahrnehmung von manipuliertem Weinen als Lachen. *Proc. 22. Konferenz Elektronische Sprachsignalverarbeitung (ESSV '11)*, Aachen, pp. 253-260.

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.

Trouvain, J. & Truong, K. 2015. Prosodic characteristics of read speech before and after treadmill running. *Proc. Interspeech*, Dresden, pp. 3700-3704.

Trouvain, J. & Möbius, B. 2018. Zu Mustern der Pausengestaltung in natürlicher und synthetischer Lesesprache. *Proc. 29th Konferenz Elektronische Sprachsignalverarbeitung (ESSV '18)*, Ulm, pp. 334-341.

Whalen, D.H., Hoequist, Ch.E. & Sheffert, S. 1995. The effects of breath sounds on the perception of synthetic speech. *Journal of the Acoustical Society of America* 97, pp. 3147-315.

Winter, B. & Grawunder, S. 2012. The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics* 40, pp. 808-815.

Włodarczak, M. & Heldner, M. 2017. Respiratory constraints in verbal and non-verbal communication. *Frontiers in Psychology* 8, article id 708.

Hearing smiles and smiling back

Pablo Arias
STMS UMR9912
(IRCAM/CNRS/Sorbonne-U)
arias@ircam.fr

Pascal Belin
INT UMR7289
(CNRS/Aix-Marseille-U)

Jean-Julien Aucouturier
STMS UMR9912
(IRCAM/CNRS/Sorbonne-U)

Abstract

Smiling may be one of the most important gestures in the human emotional repertoire. Long thought to be a universal expression of positive affect and affiliation, recent findings suggest the smile gesture is a highly adaptive, functionally diverse, and culturally variable gesture. But, although research has shed light on smiles as a visual communicative behavior, it is less known that smiles also have consequences in another modality : audition. As a gesture which changes the shape of the main vocal resonator (the mouth), smiling while speaking acts as an acoustic filter, creating smile-related acoustic structures in the sound. But what are these acoustic structures? And are the cognitive mechanisms involved in processing such auditory cues similar to those involved in processing visual smiles? Here, we briefly report on a series of studies addressing these questions. First, we describe the acoustic fingerprint of auditory smiles as measured by reverse correlation, then present a computational model to parametrically control smile specific acoustic cues in speech, and finally present facial electromyography data suggesting that these acoustic cues are not only recognized by naive participants but can also trigger low-level imitative mechanisms that are usually associated with the processing of visual emotions.

1 Introduction

Since the seminal work of Darwin and Duchesne, the expressive function of smiles have intrigued scientists across disciplines (Darwin, 1872). Step after step, research has found evidence of their early development (Meltzoff and Moore, 1977; Oostenbroek et al., 2016), their presence across cultures (Ekman et al., 1969; Jack et al., 2012), their use in other species (Parr and Waller, 2006) and their functional diversity (Rychlowska et al., 2017).

The neural processing of such facial expressions is complex. Recently, a framework based on embodied simulations was suggested to understand the underlying mechanisms necessary to the processing of a smile (Niedenthal et al., 2010). In such a framework, the perception of the smiles is thought to activate motor/articulatory mechanisms leading the observer to simulate, in their own body, the facial expression of the sender.

Yet, interestingly, smiles do not only have visual consequences. It is an open secret that smiles can also be heard in speech, even in the absence of visual cues (Tartter, 1980; Basso and Oullier, 2010). In a source-filter perspective, stretching lips while speaking changes the shape of the vocal resonator, possibly reducing vocal tract length, and thus transmitting filtered frequency content from the glottal impulses when compared to normal speech. But does the auditory perception of a smile also trigger the embodied mechanisms we associate with their visual counterparts? Do auditory smiles also trigger congruent facial activity when perceiving them?

In this work we will present a recent series of studies (Ponsot et al., 2018; Arias et al., 2018a,b) describing first how auditory smiles are acoustically defined, and second, how their processing can trigger facial mimicry, which suggests that embodied mechanisms are also present and used to process smiles aurally, shedding light on the amodal nature of such emotional/articulatory processing.

2 Reverse correlating smile's acoustic fingerprint

In a first study (Ponsot et al., 2018), we aimed to characterize what an auditory smile is from an auditory perspective. Initially, auditory smiles or smiled speech were thought to involve similar prosody to that of happy speech, with high mean pitch and high intensity (Quené et al., 2012; Barthel and Quené, 2015; Lasarcyk and Trouvain, 2008). However, because smiles can also be perceived in whispered, non-pitched voices (Tartter and Braun, 1994), pitch and prosody do not appear necessary components of smiled speech, which may more primarily affect sound spectrum. Accordingly, smiled speech was found associated with an increase of certain formant frequencies and amplitude (Barthel and Quené, 2015; Podesva et al., 2015; El Haddad et al., 2015, 2017), as well as F1-F2 dispersion (Drahota et al., 2008).

To complement these results, we conducted a reverse correlation experiment in order to study the perception as opposed to the production — of auditory smiles. N=10 participants were presented hundreds of pairs of [a] phonemes utterances which had randomly manipulated spectral characteristics and were asked to indicate, in each pair, which was the most smiling. We then used psychophysical reverse correlation (Ahumada Jr and Lovell, 1971) to derive the mean filter needed to transform the spectral properties of the target phoneme in order to be recognized as 'smiled'. The filter presented finely-tuned structures around the original phoneme's formants, implementing a formant shift of both F1 and F2, and an amplitude boost for F3 (Fig. 1; see (Ponsot et al., 2018) for details).

3 Modeling auditory smiles

To study how auditory smiles are perceived on arbitrary speech, we then designed a digital audio transformation algorithm able to simulate the acoustic changes seen in the above reverse correlation experiment (Arias et al., 2018b). The aim of the algorithm was to operate on real speech, changing only smile specific acoustic features while preserving its identity, prosody and content. Therefore, we implemented a transformation of the spectral envelope, which preserves the harmonic partials of the original voice, avoiding artifacts caused by the synthetic glottal impulses. The algorithm shifts the two first formants of the voice, and then applies a boost of the third formant. The algorithm has two modes : the smile, in which formants are shifted towards the high frequencies, and the unsmile mode, in which formants are shifted towards the low frequencies. In several validation experiments, we found the manipulation significantly affected listeners' impression of speaker's smiliness in arbitrary spoken sentences, as well as ratings of speaker emotions that involved stretched/contracted lips (e.g., *joy* or *irony*) while having no effect on ratings of emotions involving other mouth shapes, such as anger or surprise (see (Arias et al., 2018a) and (Arias et al., 2018b) for details) .

4 Auditory smiles trigger unconscious facial imitation

Finally, to study the mechanisms underlying the processing of auditory smiles, and probe their similarity with their visual counterparts, we designed an electromyography (EMG) experiment aiming at measuring facial reactions while listening to smile-related acoustic cues (Arias et al., 2018a). We first transformed a set of 20 sentences with neutral content both with the smile and the unsmile audio effects and asked N=35 participants to judge their smiliness. We recorded participants' Zygomaticus major muscle (involved in smiling), and Corrugator supercili (involved in frowning) while participants were listening to the sounds.

Participants' rated smile transformed sentences as being more smiling, and their facial reactions were congruent with their judgements, with significant differences between smile and unsmile time courses of EMG activity for both Zygomatic (cluster permutation test: t=1.1-1.9sec.; p=0.001; d=0.52) and Corrugator (t=0.8-1.6sec.; p=0.008; d=-0.41; figure 1-b).

A more detailed analysis with GLMMs (General Linear Mixed Models) revealed a functional dissociation between the two muscles. For the zygomatic muscle, there was a main effect of sound manipulation (smile/unsmile; $\chi^2(7)$=6.6, p=0.01) and participants' rating ($\chi^2(7)$=4.5, p=0.03) whereas for the corrugator muscle, there was a main effect of rating ($\chi^2(7)$= 27.2, p=1.7e-7) but no effect of sound manipulation ($\chi^2(7)$=0.88, p=0.35). In other words, imitative behavior on the corrugator muscles was entirely mediated by participants' ratings, whereas zygomatic activity was also independently driven by the low-level acous-

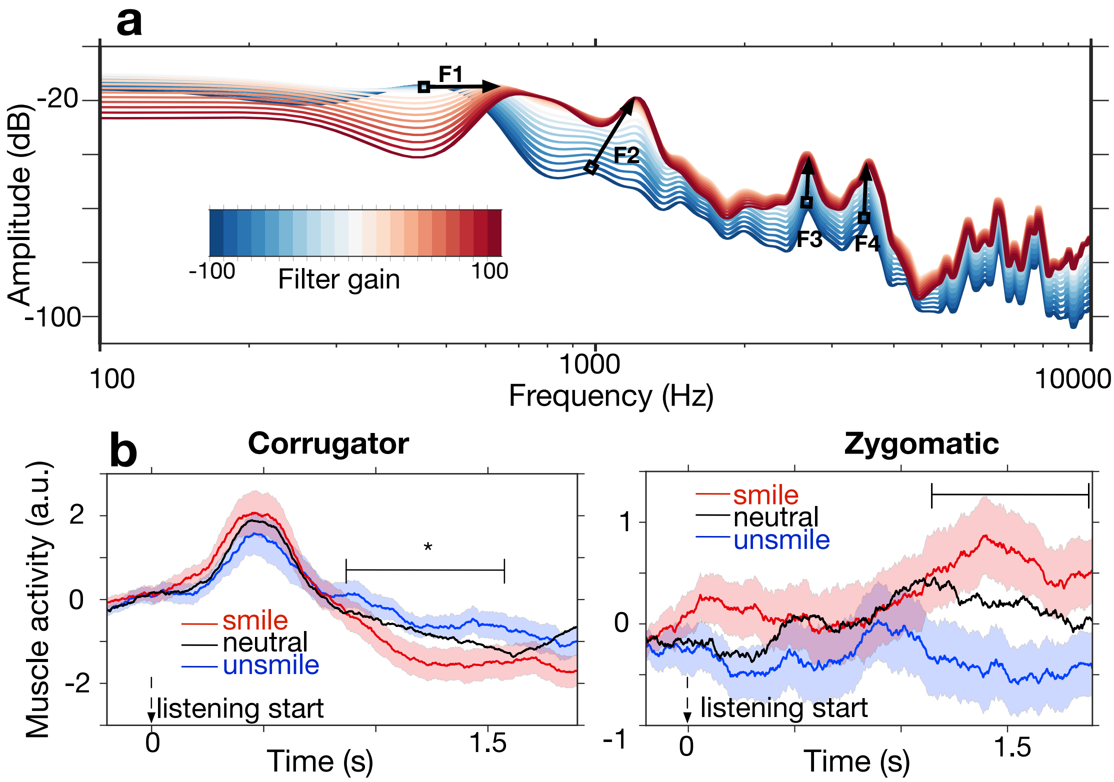


Figure 1: (a) Mean spectral envelope across participants when applying different gains to the mean filter found in the reverse correlation experiment, highlighting the overall transformation over the spectral envelopes as one goes from a strongly non-smiling voice (blue), to that of a strongly smiling voice (Red). (b) Participants' corrugator (left) and zygomatic (right) EMG activity while rating speaker smiliness for neutral (black), smile- (red) and unsmile-transformed (blue) versions of 20 sentence stimuli, displayed as a function of time. Asterisks indicate time intervals showing statistically significant differences between smile and unsmile conditions assessed with cluster permutation tests;

tic cues manipulated by the algorithm, even when participants did not detect them (see (Arias et al., 2018a) for details).

5 Discussion

In this paper, we've reported on a recent series of studies we conducted to shed light on the representation and cognitive processing of auditory smiles. First, we characterize auditory smiles' spectral structures, which are defined by specific formant movements both in amplitude and in frequency. Second, we presented data suggesting that these acoustic cues can trigger rapid and congruent facial reactions. These results provide evidence of auditory-based facial mimicry when processing the acoustic cues caused by smiling in speech. These results suggest that the embodied mechanisms often associated with the visual processing of facial expressions can also be activated acoustically as long as the sound also conveys articulatory/emotional information. More generally, these results suggest that there may be shared, or at least similar, mechanisms to process both visual and auditory smiles.

Acknowledgments

Work funded by ERC StG CREAM 335536 and ANR 2017 REFLETS to Jean-Julien Aucouturier.

References

Al Ahumada Jr and John Lovell. 1971. Stimulus features in signal detection. *The Journal of the Acoustical Society of America*, 49(6B):1751–1756.

Pablo Arias, Pascal Belin, and Jean-Julien Aucouturier. 2018a. Auditory smiles trigger unconscious facial imitation. *in press*.

Pablo Arias, Catherine Soladie, Oussema Bouafif, Axel Robel, Renaud Seguier, and Jean-Julien Aucouturier. 2018b. Realistic transformation of facial and vocal smiles in real-time audiovisual streams. *IEEE Transactions on Affective Computing*.

Helen Barthel and Hugo Quené. 2015. Acoustic-phonetic properties of smiling revised–measurements on a natural video corpus. In *Proceedings of the 18th International Congress of Phonetic Sciences.–Glasgow, UK: The University of Glasgow*.

Frédéric Basso and Olivier Oullier. 2010. smile down the phone: Extending the effects of smiles to vocal social interactions. *Behavioral and Brain Sciences*, 33(06):435–436.

Charles Darwin. 1872. *The expression of the emotions in man and animals*. Oxford University Press, USA (1998).

Amy Drahota, Alan Costall, and Vasudevi Reddy. 2008. The vocal communication of different kinds of smile. *Speech Communication*, 50(4):278–287.

Paul Ekman, E Richard Sorenson, and Wallace V Friesen. 1969. Pan-cultural elements in facial displays of emotion. *Science*, 164(3875):86–88.

Kevin El Haddad, Stéphane Dupont, Nicolas d'Alessandro, and Thierry Dutoit. 2015. An hmm-based speech-smile synthesis system: An approach for amusement synthesis. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, volume 5, pages 1–6. IEEE.

Kevin El Haddad, Ilaria Torre, Emer Gilmartin, Hüseyin Çakmak, Stéphane Dupont, Thierry Dutoit, and Nick Campbell. 2017. Introducing amus: The amused speech database. In *International Conference on Statistical Language and Speech Processing*, pages 229–240. Springer.

Rachael E Jack, Oliver GB Garrod, Hui Yu, Roberto Caldara, and Philippe G Schyns. 2012. Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109(19):7241–7244.

Eva Lasarcyk and Jürgen Trouvain. 2008. Spread lips+ raised larynx+ higher f0= smiled speech?-an articulatory synthesis approach. *Proceedings of ISSP*, pages 43–48.

A.N. Meltzoff and M.K. Moore. 1977. Imitation of facial and manual gestures by human neonates. *Science*, 198:7578.

Paula M Niedenthal, Martial Mermillod, Marcus Maringer, and Ursula Hess. 2010. The simulation of smiles (sims) model: Embodied simulation and the meaning of facial expression. *Behavioral and brain sciences*, 33(6):417–433.

Janine Oostenbroek, Thomas Suddendorf, Mark Nielsen, Jonathan Redshaw, Siobhan Kennedy-Costantini, Jacqueline Davis, Sally Clark, and Virginia Slaughter. 2016. Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology*, 26(10):1334–1338.

Lisa A Parr and Bridget M Waller. 2006. Understanding chimpanzee facial expression: insights into the evolution of communication. *Social Cognitive and Affective Neuroscience*, 1(3):221–228.

Robert J Podesva, Patrick Callier, Rob Voigt, and Dan Jurafsky. 2015. The connection between smiling and goat fronting: Embodied affect in sociophonetic variation. In *Proceedings of the International Congress of Phonetic Sciences*, volume 18.

Emmanuel Ponsot, Pablo Arias, and Jean-Julien Aucouturier. 2018. Uncovering mental representations of smiled speech using reverse correlation. *The Journal of the Acoustical Society of America*, 143(1):EL19–EL24.

Hugo Quené, Gün R Semin, and Francesco Foroni. 2012. Audible smiles and frowns affect speech comprehension. *Speech Communication*, 54(7):917–922.

Magdalena Rychlowska, Rachael E Jack, Oliver GB Garrod, Philippe G Schyns, Jared D Martin, and Paula M Niedenthal. 2017. Functional smiles: Tools for love, sympathy, and war. *Psychological science*, 28(9):1259–1270.

Vivien C Tartter. 1980. Happy talk: Perceptual and acoustic effects of smiling on speech. *Attention, Perception, & Psychophysics*, 27(1):24–27.

Vivien C Tartter and David Braun. 1994. Hearing smiles and frowns in normal and whisper registers. *The Journal of the Acoustical Society of America*, 96(4):2101–2107.

Clarifying Laughter

Chiara Mazzocconi[1], **Vlad Maraev**[2], **& Jonathan Ginzburg**[1,3]
[1]Laboratoire Linguistique Formelle (UMR 7110)
& [2]Centre for Linguistic Theory
and Studies in Probability (CLASP),
Gothenburg University & [3]Laboratoire d'Excellence (LabEx)—EFL
Université Paris-Diderot, Paris, France
tiany.03@gmail.com

Abstract

In the current paper we investigate whether laughter can be object of clarification requests and what these clarification requests might be about. We use the range of possible clarification requests as diagnostics for the constitutive elements of the meaning conveyed, thereby testing existing hypotheses concerning the semantics of laughter.

1 Introduction

Although laughter has been of interest to philosophers for millennia and in recent times studied extensively by psychologists, neuroscientists, and phoneticians, it has been assumed to lack propositional content (Hepburn and Varney, 2013). Ginzburg et al. (2015) provide extensive evidence to the contrary, on the basis of its stand alone uses as a response or follow up to questions and assertions, and its intra-utterance use to effect scare quoting. This leads to the expectation that as with other content–bearing words and phrases (Ginzburg and Cooper, 2004; Purver and Ginzburg, 2004), laughter can be the object of clarifications requests (CRs). In this paper, the first to our knowledge to broach this issue, we show that this expectation is met and we use the range of potential clarifications as diagnostics to identify some of the constituents of laughter meaning.

In section 2 we present some previous studies about laughter which lead to the current investigation, in section 3 we present and analyse some examples, sources and forms of clarification requests and of spontaneous clarifications. Finally, section 4 is a brief discussion and conclusion section.

2 Background

Ginzburg et al. (2015) and Mazzocconi et al. (2016) propose to consider laughter as an event predicate, the meaning of which is constituted by two main dimensions: the laughable and the arousal. By laughable we mean the argument the laughter predicates about. Different kinds of laughable can be distinguished firstly based on whether they contain an incongruity or not and secondly depending on which kind of incongruity it is, being therefore a categorical variable. Arousal, intended as intensity of activation/wakefulness, on the contrary is a continuous one: going from very low (e.g. little giggle, quiet laughter) to very high (e.g. loud uncontrollable laughter). In the framework adopted incongruity is defined, as proposed in Ginzburg et al. (2015), as a clash between a general inference rule (a topos) and a localized inference (an enthymeme), a view inspired by work in humour studies e.g., Hempelmann and Attardo (2011). To exemplify: (1a) is an enthymeme, an instance of the topos in 1b). A's utterance (3) in (1c) relies on the enthymeme in (1d), which clashes with the topos in 1b). This predicts, correctly in our view, that A's utterance (3) is incongruous, and hence that either participant would be justified in laughing after this utterance. Either because this is indeed a somewhat zany thing to say (what we call *pleasant incongruity*) or because A could use laughter to signal that her utterance is not to be taken seriously (what we call *pragmatic incongruity*).

(1) a. Given that the route via Walnut street is shorter than the route via Alma, choose Walnut street.

b. Given two routes choose the shortest one.

c. A(1): Which route should I choose?
B(2): The route via Walnut street is shorter.
A(3): OK, so I will choose the route via Alma.

Search	SWBD	CRs	BNC	CRs
Laughter occurrences	26861		30598	
What's funny	5		5	4
What's so funny	3		3	1
What are you laughing about	0		0	
What are you laughing at	0		2	2
What you laughing for	0		2	2
Why are you laughing	0		0	
That's not funny	1		4	
Why do you find that funny	0		0	
Do you find that funny	0		0	
Why do you laugh	0		0	
What are you laughing at	0		2	2
What's that loud laughter	0		0	
What's that laugh	0		0	
Why so loud	0		0	
Laugh because	7		2	
Laughing at	4		55	

Table 1: Results search in Score: SWBD and BNC data.

d. Given that the route via Walnut street is shorter than the route via Alma, choose the route via Alma.

We propose, following Ginzburg et al. (2015) and Mazzocconi et al. (2016), that the core meaning of laughter involves a predication $P(l)$, where P is a predicate that relates to either *incongruity* or *closeness* and l is the laughable, an event or state referred to by an utterance or exophorically. As Ginzburg et al. (2015) show, this core meaning, when aligned with rich contextual reasoning, can yield a wide range of functions.

(2) Laughter meaning: The laughable l having property P triggers a positive shift of arousal of value d within A's emotional state e.

3 Clarification Request Data

The data analysed are taken from 2 different corpora: the British National Corpus (BNC) (Burnard, 2000) and the Switchboard corpus (SWBD) (Godfrey et al., 1992), searched using the SCoRE search engine (Purver, 2001).

Despite the very high number of laughter occurrences (Table 1) in both corpora (26,861 in SWBD and 30,598 in BNC), we found very few explicit CRs for laughter (0 in SWBD and 11 in BNC; 0.03 % of all the laughs produced). This frequency is significantly smaller than that found for nominals in Purver (2004) (46 CRs over a total of 24,310 common nouns produced (0.18%)), but is of a similar order to the frequency found for verbs (3 CRs over a total of 30,060 verb occurrences (0.09%)).[1] One does, nonetheless, find regular occurrences of participants spontaneously providing explicit justifications of their laughter behaviour to make sure the interlocutors interpret correctly their contribution, providing information about the elements necessary for a laughter to occur.

3.1 Sources

The first question of our interest was which were the causes of problematic interpretation of a laughter and, maybe not surprisingly, we found that the element asked to be clarified most frequently is the laughable, i.e. the argument of the laughter predication.

3.1.1 Laughable

Our data show that the highest number of CRs related to laughter assume as default that the predication involves funniness i.e. predication of the presence of a pleasant incongruity in the laughable, which could be paraphrased as "This is funny!". Therefore typical CRs related to a laugh are "What's funny?" "What's so funny?". This can be explained given data from Mazzocconi et al. (2016) that shows a high frequency of laughter predicating about pleasant incongruities used to show enjoyment of those, in comparison to the other types of laughables and functions; this is consistent also with the fact that this use of laughter is the more ancient and basic one both phylogenetically and ontogenetically.

1. **Argument - pleasant incongruity**: In (3) the CR about the argument of the laughter is met by pointing at what Mazzocconi et al. (2016) classify as a metalinguistic laughable (e.g., a slip of the tongue, pun, violation of conversational rules, inappropriate speech act etc.). This relates not to the content of Andrew's utterance, but to its form. While in (4) the laughable is clarified by describing verbally the gossip considered to be funny by Daniel and the Unknown speaker.

(3) *Extract from BNC, KBW*
Tim: I don't want chocolate.
Dorothy: Shh. Shh.< unclear >
Andrew: Tim. If you don't want to finish it just put it down there and keep quiet.
Dorothy: < laugh >
Andrew: **What are you laughing at?**
Dorothy: < laughing > the way you said it .

(4) *Extract from BNC, KNY*
Unknown: < laugh >
Marc: What was that you said?
Alex: Nothing.
Marc: James, **who's he laughing at?** What have you been saying?

[1] An explanation of the noun/verb differences is still elusive (anon2, 2017).

Emma: James.
Unknown: Alex please < unclear >.
Daniel: James[last or full name]fancies Zoe.
Emma: Does he?

2. **Argument - retracting funniness assumption**: In (5) it seems that the default interpretation of the laughter production "my partner has perceived something funny", justifies the question "what's funny?"; when the expected answer is not provided, this is then retracted in "What are you laughing at then?", Angela becoming open to the other possible laughter functions and laughable types.

(5) *Extract from BNC, KSS*
Angela: **What's funny?** < pause > What you doing?
Richard: I'm not doing a thing. You're doing it.
Angela: **What you laughing at then?**
Arthur: < unclear >.< laugh >
Angela: You're waiting for what? What you waiting for?

3. **Argument - pragmatic incongruity** We did not find CRs related to pragmatic incongruity (i.e. when there is a clash between what is said and what is intended). However, this absence, we think, can be explained by the scarcity of this kind of laughable (in anon (subm) over 1072 laughter only 1% were related to a pragmatic incongruity). We can indeed imagine contexts in which a CR for this type of laughable could be quite natural:

(6) *Constructed example*
A: She is Johns long-term, < laughter/ > friend.
B: < laughter/ > **Why that snigger?** < laughter/ > Is there something more than friendship?

4. **Topoi and enthymemes**: In (7) and (8) the person asking for clarification does not have any issues identifying the laughable in itself, it is very clear for them what the interlocutor is *laughing about*; the objects of their CRs are, we argue, the topos and the enthymeme implicated in the incongruity. In (7) Geoff probably understood which topos and enthymeme his mum is considering, but still asks for elaboration. In (8) Anon explains very clearly the reason for his/her pleasant incongruity appraisal stating that he wouldn't expect "him to do that", therefore pointing at a clash between expectations and reality.

(7) *Extract from BNC, KD6*
Geoff: ah
Lynn: < laugh/ >
Geoff: I like that
Lynn: gosh
Geoff: **What you laughing for?**, **I wouldn't laugh**
Lynn: oh
Geoff: silly mummy < pause > oh dear table's wobbling

(8) *Extract from BNC KST*
Margaret: Yes, but pretend she's not watching and he looks over the top of his paper.
Anonymous: And grins!
Margaret: Oh it's stupid! I mean if anybody else just got up on the stage like he does < pause > and kicks his leg, kick like their leg like er like that they'd boo him off!
Anonymous: It's quite funny though < pause > when he kicks his legs and he went< unclear >he goes< pause >ooh wah!
Margaret: **What's funny about it**?
Anonymous: **Well that's funny! You're not expecting him to do that.**

3.1.2 Arousal

The second laughter dimension proposed in Mazzocconi et al. (2016) is arousal. There are two things that can be questioned about the shift in arousal a laughter signals: the direction (i.e. positive – enjoyment) and the amplitude of such shift. In example 9 Danny asks a CR about the enjoyment (positive shift in arousal) felt by Mark inferred from his laughter. On the other hand it is possible for a CR to be posed when the arousal perceived clashes with our evaluation of the laughable, questioning therefore the amplitude of the shift. We can imagine a situation as in (10), in which A is puzzled about the extremely highly aroused laughter produced by B when looking at the vignette s/he is showing her and in asking for clarification s/he is implicitly asking for the topos and enthymeme utilised, because according to the ones A considered such aroused laughter would be inappropriate.

(9) *Extract from BNC, F7U*
Danny: < pause > Yes, that's what it means, it means weighing scales. < pause > What he meant was a balance.
Mark: < laughter/ >
Danny: Erm < pause > right if this < pause > < laughter/ > **you're enjoying this Mark aren't you?** < pause > Dunno why, they'll start me off now!

(10) *Constructed example*
A: Look at this vignette! Isn't it nice? < laughter/ > [=little giggle]
B: < laughter/ > < laughter/ > [=bursting out laughing very loudly and uncontrollably]
A: **What's that loud laughter???**
B: < laughter > It made me think about what happened that day with my friend... < laughter/ > etc.

3.2 Form

The second aspect of our interest is the form CRs related to laughter can have. With nouns and verbs it is indeed possible to ask for clarification in different ways: from full sentences which echo or reprise the source; via non sentential, elliptical fragments containing only noun phrases or wh-phrases; to highly conventionalised particles like "Eh?" (Purver, 2004). Based on our corpus analysis it appears that not all of these forms are viable when asking for laughter clarification.

1. **Direct CRs**
   In our exploration most of the direct CRs we could find were wh-phrases (see examples above 3, 4, 5, 7, 8) directed either at the argument or the arousal of the laughter produced. While in 9 we have a confirmation clausal question (Ginzburg and Cooper, 2004).

2. **Echoing-reprising the source**
   We can nevertheless imagine other contexts in which a reprise (or a non-reprise (Purver, 2004)) of the source is used to construct a CR.

   (11) *Constructed example*
   A: So you know... now there are gonna be important political consequences after yesterday's demonstration.
   B: < laughter/ >
   A: **Ahah?? / What do you mean "Ah-ah"?? / "Ah-ah" What?**
   B: Well, you know! Do you really expect something good?? What are they gonna do! As usual some useless declaration on tv and that's all.

   A consideration to be stressed is though that the latter kinds of CR would probably work best only with low arousal laughter with enough harmonic elements, given the need for modulating the prosodic contour into a question-like intonation. Therefore a question here arises about whether different kinds of laughter allow different forms of CRs.

3. **Indirect CRs**
   It is also possible to use very indirect ways of asking for clarification which are much harder to spot in a large corpus. Here is an example from the St. Louis Post-Dispatch:

   (12) *Example from St. Louis Post-Dispatch - 11 May 2018*
   The defense objected and Burlison sustained the objection. Sullivan laughed.
   "Is there something about my ruling that strikes your fancy?" Burlison said.
   "No," Sullivan replied, "Im laughing to myself about something else."

4 Discussion and Conclusion

The data presented raises a variety of questions. We mention briefly two: first: why are few occurrences of laughter CRs found? Second: why are they all related to laughs concerning pleasant incongruities and none concerning social, pragmatic incongruities or closeness. The answer to these questions might be correlated. On the one hand it is possible that a more refined exploration of the corpus will allow the detection of more indirect forms of CRs. On the other hand we think that a laughter CR is potentially rude or aggressive. That might explain, given its exclusive reliance on phone conversations between strangers, why in SWBD we do not find any direct laughter CR. Issues related to politeness and social conventions might also explain the absence of laughter CRs related to social incongruities. In these kind of situations the request for a clarification would indeed have a contrary effect to the one aimed by the laugher, making the situation very uncomfortable for the parties involved. These kinds of laughter are moreover usually very low arousal and people are often not even aware of producing them (Vettin and Todt, 2004).

References

anon. subm. What's your laughter doing there? A taxonomy of the pragmatic functions of laughter. *(Under Review).*

anon2. 2017. Lexical categories and clarificational potential. Revised version under review.

Lou Burnard. 2000. Reference guide for the british national corpus (world edition).

Jonathan Ginzburg, Ellen Breitholtz, Robin Cooper, Julian Hough, and Ye Tian. 2015. Understanding laughter. In *Proceedings of the 20th Amsterdam Colloquium*, University of Amsterdam.

Jonathan Ginzburg and Robin Cooper. 2004. Clarification, ellipsis, and the nature of contextual updates in dialogue. *Linguistics and philosophy*, 27(3):297–365.

John J Godfrey, Edward C Holliman, and Jane McDaniel. 1992. Switchboard: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, volume 1, pages 517–520. IEEE.

Christian F Hempelmann and Salvatore Attardo. 2011. Resolutions and their incongruities: Further thoughts on logical mechanisms. *Humor-*

International Journal of Humor Research, 24(2):125–149.

Alexa Hepburn and Scott Varney. 2013. Beyond ((laughter)): some notes on transcription. In Philip Glenn and Elizabeth Holt, editors, *Studies of Laughter in Interaction*. Bloomsbury.

Chiara Mazzocconi, Ye Tian, and Jonathan Ginzburg. 2016. Towards a multi-layered analysis of laughter. In *Proceedings of JerSem, the 20th Workshop on the Semantics and Pragmatics of Dialogue (SemDial), Rutgers, the State University of New Jersey, New Brunswick, USA*.

Matthew Purver. 2001. SCoRE: A tool for searching the BNC.

Matthew Purver. 2004. *The theory and use of clarification requests in dialogue: Kings college, University of London Ph. D.* Ph.D. thesis, dissertation.

Matthew Purver and Jonathan Ginzburg. 2004. Clarifying noun phrase semantics. *Journal of Semantics*, 21(3):283–339.

Julia Vettin and Dietmar Todt. 2004. Laughter in conversation: Features of occurrence and acoustic structure. *Journal of Nonverbal Behavior*, 28(2):93–115.

Laugh is in the air? Physiological analysis of laughter as a correlate of attraction during speed dating

Susanne Fuchs
Leibniz-Zentrum Allgemeine
Sprachwissenschaft (ZAS)
Schützenstrasse 18, 10117 Berlin
fuchs@leibniz-zas.de

Tamara Rathcke
English Language & Linguistics
University of Kent
Canterbury, Kent, CT2 7NF
T.V.Rathcke@kent.ac.uk

Abstract

Laughter has rarely been examined in the context of a romantic attraction between conversational partners. This study aims to fill the gap by investigating laughter in the context of speed dating. We present a preliminary analysis of spontaneous laughs produced by male and female speakers who were more or less attracted to each other, and discuss variability across individuals in terms of laughter frequency and overlap.

1 Introduction

Laughter is a universal non-verbal expression (Sauter et al., 2010) that has its highest occurrence in human interaction, and is not primarily related to the presence of humor (Vettin and Todt, 2004). For instance, in a current meta-analysis by Montoya et al. (2018) laughing has a clear social component, since it has been identified as one of several features related to self-reported attraction. Scott et al. (2014) also commented on the social life of laughter and differentiated between voluntary and involuntary laughter.

A strong physiological component involved in laughing is respiration. Scott et al. (2014) describe laughter as a spasm in the diaphragm and the chest wall. Filippelli et al. (2001) analyzed respiratory dynamics in laughter and found sudden decrease in lung volume that goes hand in hand with rapid decrease to residual respiratory capacities. In his analysis on respiration in human conversation, McFarland (2001) found increased synchronization in respiratory kinematics among interlocutors during laughter. Shared physiological states

Speed dating is has recently been used to elicit different degrees of inter-personal attraction and their corresponding prosodic features (Michalsky et al., 2017), and has not been studied extensively in terms of interpersonal synchrony, including laughter. The overarching aim of the study is to inform the accommodation theory, primarily by providing evidence on how interpersonal accommodation is influenced by interpersonal attraction. The results reported here focus on the laughing behavior of several heterosexual speed-daters who are less or more, mutually or unrequitedly, attracted to each other.

2 Methodology

2.1 Participants and experimental procedures

All participants were recruited via a local participant database. The call for participation in a speed dating study went out to all heterosexual, single participants, native speakers of German aged between 20 and 30 years. Two males and six females volunteered to take part. The participants were asked to come to the laboratory on two successive days. On the first day, experimental set-up and all procedures were explained, and participants had a chance to familiarize themselves with the setting and the equipment, and to ask questions. Moreover, a baseline recording was obtained when participants talked to a same-sex dialogue partner for 5 minutes, which resulted in eight baseline dialogues. In addition, breathing was recorded in quiet for 3 minutes and in several respiratory maneuvers (vital capacity, iso-volume).

The speed dating experiment took place on the following day. Both male speakers talked to six female participants, which resulted in 12 target dialogues in total. We ensured that participants would not see each other before the recording and directed them in a waiting room upon arrival (there was a separate waiting room for male and for female participants).

While waiting, participants got prepared for the recording session by putting on a motion capture headband and jacket with markers. One respiratory belt was placed around the thorax and one around the abdomen. Two assistants were present to support with this.

During each dating session, participants were placed at a table, the respiratory belts and head-mounted microphones were connected and the experimenters checked all signals as quickly as possible. Using a smartphone, participants rated each dating partner on a 10-point Likert scale (from 1 least up to 10 most attractive) before and after their 5-min dating conversation. The session was stopped after a 5-min conversation by the experimenter. Participants agreed that their contact details would only be given to another person if both parties were interested. If only one of the daters expressed interest in the other, no contact details were released.

2.2 Experimental set-up

Participants were recorded with a multimedia set-up involving a motion capture system (OptiTrack, *Motive* Version 1.9.0) with 12 cameras (Prime 13) to record upper body motions. Inductance Plethysmography (formerly Respitrace) was used to record breathing behavior simultaneously. Speech acoustics was recorded together with the motion capture system, with the Inductance Plethysmograph and as a stereo signal on DAT. At the beginning and end of each recording, synchronization impulses were send from the motion capture system to the computer where breathing signals were recorded which allowed the synchronization of the signals. In addition, acoustic data recorded on each computer involved (Inductance Plethysmograph, motion capture, DAT) were used to check the synchronization of all signals via cross-correlation. During speed dating, participants were left alone in the recording room to feel less observed. To bring some romantic mood into the lab, the testing room was decorated with green plants, posters and flowers (see Figure 1). The computers were moved outside of the lab to control the signals in real time.

2.3 Data pre-processing and annotation

The respiratory signals have been cut according to the synchronization impulses from the motion capture signal using MATLAB (R2017b). All speed dating dialogues have been transcribed by means of the acoustic signals using Praat version 6.0.37 (Boersma and Weenink, 2018). Thoracic and abdominal volume changes have been summed for each participant to consider the whole lung volume and not only parts of it. Onsets of inhalation and inhalation peaks have been determined manually in Praat. One breathing cycle was then defined from one inhalation onset to the next. Laughter has been first identified in the acoustic signal and then in the respiratory data. The beginning of laughter has been labeled as the time point where exhalation starts to drop rapidly. The end of laughter was determined as the minimum of exhalation. We defined shared laughter as the temporal overlap in laughter between two partners (an example is given in Figure 2).

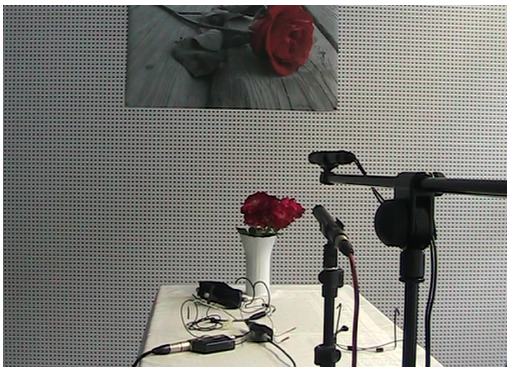


Figure 1: Laboratory with table, poster and roses.

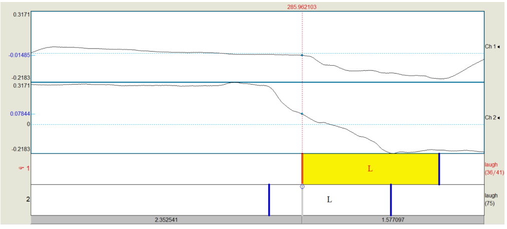


Figure 2: First and second track correspond to respiratory signals of two interlocutors. Their corresponding laughter (drop in exhalation) has been annotated below (annotation in yellow corresponds to the breathing signal in the first track and bottom annotation to the second track). Both laughter determined on the respiratory signal overlap in time.

3 Results and Discussion

Only a subset of the data has been analysed so far. The results below are based on some preliminary analyses of 5/6 dialogues for s8 and 3/6 dialogues for s1. Most of these dialogues started with several bouts of laughter, which is unusual in same-sex conversations (0-2.3% of all recording laughter bouts in Vettin and Todt 2004). Following Montoya et al. (2018), our finding may be interpreted with respect to the participants' insecurity due to not knowing each other at the beginning of the experiment and their aim to quickly develop interpersonal trust and rapport (given that their time together was strictly limited to 5 minutes).

There was one perfect match between participants (between the male speaker s1 and the female speaker s3), i.e. both participants were interested in meeting each other again after the experiment. The feeling of attraction was not shared in several pairs: female speakers s2 and s3 were interested in meeting s1 (but not vice versa). In contrast, both male participants were interested in meeting s3, s5, s6, s7, but not s2 or s4. None of the female participants expressed an interest in seeing again s8. To what extent is the interpersonal attraction reflected in the participants' laughter? Preliminary results are show in Table 1. Overall, female speakers laughed more often than their male interlocutors. There were only two conversations where both speakers laughed at a similar rate (s8 and s3), or the male speaker laughed more often (s1 when paired with s4).

There was a striking differences in laughter frequency between the two male participants: while s1 laughed similarly often with all of his partners, s8 changed his laughing behavior depending on his attraction to his female interlocutor. More specifically, s8 laughed almost twice as often when talking to s5 and s6 in comparison to the two females he was not interested in (s2 and s4). A similar pattern could be attested among the female speakers: while the laughter frequency was not influenced by the attraction in s2, s3 laughed more in her conversation with s1 (who she liked) than s8 (who she was not interested in). Interestingly, s3 and s8 who shared differential laughter frequency are the only participants of the study who had never had a long-term relationship.

Shared laughter, i.e. laughter where both partners overlap in respiratory activity and the resulting laughter bouts coincide in time, seems to be the core feature that gives away the attraction felt by s1 and s2. They both laughed more often in synchrony with partners they fancied (s3 or s1, respectively). Moreover, shared laughter in these data has its highest value in our perfect match (s1 with s3, 52.95%). The lowest amount of shared laughter (16%) was found in the dialogue where both interlocutors were not interested in each other (s8 and s4).

Table 1: Laughter statistics: speaker IDs; number of laughs by males (s8 or s1), number of laughs by females (s2-s7), sum of all laughs in the dialogue, difference in number of laughs between partners (M-F speakers), number of shared laughs as percentage of the total laughs in the dialogue.

pair	laughs M	laughs F	laughs total	diff. of laughs	relat. overlap
s8_s2	14	25	39	-11	30.77
s8_s3	12	11	23	1	26.09
s8_s4	10	15	25	-5	16.00
s8_s5	20	37	57	-17	35.09
s8_s6	27	29	56	-2	32.14
s8_s7	15	19	34	-4	23.53
s1_s2	17	27	44	-10	45.45
s1_s3	16	18	34	-2	52.94
s1_s4	16	8	24	8	50.00

Based on the preliminary analyses carried out so far, we may conclude that laughter plays an important role in signaling attraction among individuals, though different aspects of the laughter behavior (its overall frequency or the frequency of shared laughter) may have different importance for involved individuals. These conclusions are indeed very preliminary, but will be substantially elaborated on at the workshop.

Acknowledgments

We would like to thank all of our participants who were amazingly open, natural, friendly and supportive to the nervous experimenter crew throughout the experiment, Jörg Dreyer for technical support, Olivia Maky and Gediminas Schüppenhauer for logistic support, and the Leibniz association for financial support.

References

Boersma, P. and Weenink, D. (2018). Praat: doing phonetics by computer. version 6.0. 37.

Filippelli, M., Pellegrino, R., Iandelli, I., Misuri, G., Rodarte, J. R., Duranti, R., Brusasco, V., and Scano, G. (2001). Respiratory dynamics during laughter. *Journal of Applied Physiology*, 90(4):1441–1446.

McFarland, D. H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research*, 44(1):128–143.

Michalsky, J., Schoormann, H., and Niebuhr, O. (2017). Turn transitions as salient places for social signals–local prosodic en-trainment as a cue to perceived attractiveness and likability. *Proceedings of the Conference on Phonetics and Phonology of German-speaking areas (PundP) Berlin*, page 125.

Montoya, R. M., Kershaw, C., and Prosser, J. L. (2018). A meta-analytic investigation of the relation between interpersonal attraction and enacted behavior. *Psychological bulletin*.

Sauter, D. A., Eisner, F., Ekman, P., and Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, 107(6):2408–2412.

Scott, S. K., Lavan, N., Chen, S., and McGettigan, C. (2014). The social life of laughter. *Trends in cognitive sciences*, 18(12):618–620.

Vettin, J. and Todt, D. (2004). Laughter in conversation: Features of occurrence and acoustic structure. *Journal of Nonverbal Behavior*, 28(2):93–115.

The Underdetermined Nature of Laughter

Gary McKeown

Queens University of Belfast

Laughter is much more than a social signal response to a humorous event. It is a crucial element of social interaction and serves multiple functions. In this talk I will argue that laughter at its core is a social bonding signal. However, it is a special social signal as it achieves many of its functions by being an ambiguous with respect to the content of the communication. It is underdetermined in the language of linguistic pragmatics, and its interpretation depends on the communicative context in which it appears.

Automatic Prediction of Affective Laughter from Audiovisual Data

Reshmashree B Kantharaju
ISIR,
Sorbonne Université,
Paris, France
bangalore_kantharaju
@isir.upmc.fr

Fabien Ringeval
LIG, UGA
CNRS, Grenoble INP,
Grenoble, France
fabien.ringeval
@imag.fr

Laurent Besacier
LIG, UGA
CNRS, Grenoble INP,
Grenoble, France
laurent.besacier
@imag.fr

Abstract

In this contribution, we provide insights on emotional laughter by extensive evaluations carried out on RECOLA corpus of dyadic spontaneous interactions, annotated with dimensional labels of emotion (arousal and valence). We evaluate, by automatic recognition experiments and correlation based analysis, how different categories of laughter, such as *unvoiced laughter*, *voiced laughter*, *speech laughter*, and *speech* (non-laughter) can be differentiated from audiovisual features, and to which extent they might convey different emotions.

1 Introduction

Laughter frequently occurs during social interactions and serves as an expressive-communicative social signal. The most commonly annotated categories of laughter are voiced and unvoiced laughter (Bachorowski et al., 2001; Petridis et al., 2013); voiced laughter is rhythmical and perceived to be more positive than unvoiced laughter which is short and does not have a rhythmic pattern (Bachorowski et al., 2001). Speech and laughter overlap frequently during conversations and this is considered as another category termed, speech laughter (Trouvain, 2003). Contextualisation of laughter with respect to the expressed emotion is important, because laughter can be produced in a large variety of context, and therefore convey specific non-verbal messages. A review of literature shows that laughter can actually express various emotions like, joy, amusement, surprise, but also nervousness, embarrassment, contempt. Whereas several studies have reported on the spectrum of emotions that laughter conveys (Bachorowski and Owren, 2001; Devillers and Vidrascu, 2007), there has been no investigations – to the best of our knowledge – on the automatic recognition of the emotions conveyed by natural expressions of laughter produced during spontaneous interactions.

In order to investigate automatic recognition of emotional laughter from spontaneous multimodal data, we performed annotation of different types of laughter on the RECOLA dataset (Ringeval et al., 2013). A total of 974 instances has been annotated, and will be made publicly available[1] to the community. Further, we conduct experiments to distinguish different categories of laughter, quantify their predictive power for emotion recognition, and the impact of the context.

2 Related Work

There exists many databases which includes annotation of laughter from audiovisual data, but very few are publicly available, with only few of them specifically dedicated to laughter. The AVLC database provides 1066 audiovisual recordings of induced and posed laughter elicited from 24 subjects, while watching a funny 10 minutes stitched clip of short videos (Urbain et al., 2010). Belfast storytelling database provides 2336 laughter instances from naturalistic multimodal data associated with social interactions (McKeown et al., 2015). The BINED database has 3 sets of audiovisual recordings of emotion elicited from watching video clip (e. g., amusement) and actively engaging participants in series of tasks to induce emotions (e. g., fear, disgust, surprise, frustration) (Sneddon et al., 2012). The first set has been included in the ILHAIRE database and it consists of 565 clips of 113 participants from which 289 instances of laughter were extracted. MMLI is a multimodal database of laughter with

[1] https://diuf.unifr.ch/diva/recola/

full body movements, facial tracking, audiovisual and physiological data (Niewiadomski et al., 2013). The database is annotated as, Laughter event (time interval in which at least one of the participants laughs) and Laughter episode (Single laugh generated by one participant); one laughter event can thus be composed of several laughter episodes. In total 439 laughter events were annotated. The SEMAINE database consists of audiovisual recordings of users interacting with limited agents, which present different personalities (McKeown et al., 2012). In total 443 instances have been annotated from 345 clips from 28 participants. The SEMAINE-SAL database includes time- and value-continuous annotations of emotional dimensions (arousal and valence), but the annotators are not consistent over the recordings. The MAHNOB database includes multimodal recordings (audio, video and physiological signals) of 22 subjects from 12 different countries watching series of video clips (Petridis et al., 2013). 563 instances of laughter were collected in total, along with 51 instances of posed laughs, 67 instances of speech-laughs and 845 instances of speech, from 180 sessions.

Even though publicly available databases of laughter provide multimodal, multilingual data, with (partially) rich annotations, there is no ratings of affective behaviour available with respect to laughter, except for the SEMAINE database. Another drawback of most existing databases is the lack of spontaneous laughter from natural interactions. Although Belfast Storytelling database provides naturalistic conversational laughter the annotations are minimal and does not provide any further details on laughter categories. Therefore, for this study, we decided to perform annotation of laughter on an existing database of spontaneous socio-affective behaviors.

3 Annotation of laughter

The REmote COLlaborative and Affective interactions (RECOLA) data set is a multimodal database of spontaneous interactions in French which consists of audio, visual, electro-cardiogram (ECG) and electro-dermal (EDA) data recorded continuously and synchronously (Ringeval et al., 2013). Spontaneous interactions from 53 participants were recorded while solving a collaborative task: ”Winter Survival Task” as dyadic teams. Affective behaviour expressed by the participants was annotated with time- and value-continuous emotional dimensions (arousal and valence) by six French-speaking assistants, for the first five minutes of each recordings, and for 46 participants. A web-based annotation tool, ANNEMO was used to perform emotion ratings and the annotations were done separately for each emotional dimension.

The initial annotations of laughter was done manually using the Audacity software. The laughter instances (excluding speech laughter) were further categorised as *voiced* or *unvoiced*. In this study, we use the voicing probability and unvoiced frame ratio to automatically categorise voiced and unvoiced laughter. We computed the voicing probability of each frame using the openSMILE acoustic feature extraction toolkit (Eyben et al., 2013). In total 53 audio files were annotated with laughter episodes labeled as ‘VL’ for *voiced laughter*, ‘UL’ for *unvoiced laughter*, ‘SL’ for *speech laughter* and non laughter speech segments were labeled as ‘S’. The data is divided into two sets namely, *Unsegmented set* and *Segmented set*. The segmented set provides annotations of interactive laughter and emotional ratings across the two affect dimensions and consists of 289 instances of laughter and 1619 instances of speech. The unsegmented set only provides annotations of laughter and consists of 974 instances of laughter.

Table 1: Number of laughter episodes annotated on the RECOLA database from complete unsegmented and segmented (initial 5 minutes) audio files with emotion ratings.

Type	Unsegmented	Segmented
Unvoiced Laughter (UL)	590	159
Voiced Laughter (VL)	187	62
Speech Laughter (SL)	197	68
All Laughter (AL)	**974**	**289**

4 Experiments and Results

We extracted three acoustic feature sets, MFCC, GeMAPS, eGeMAPS using OpenSMILE toolkit (Eyben et al., 2013) and a visual feature set based on 17 facial action units using OpenFace (Baltrušaitis et al., 2015). We further combine the best performing audio feature set with video (early fusion). For both classification and regression tasks, we make use of LIBLINEAR, an open source library for large-scale linear classification. To take into account speaker depen-

Table 2: Results (%UAR) for 2 class, 3 class and 4 class classification tasks using different feature sets; best results over the feature sets are highlighted in bold.

Type	Task	MFCCs	GeMAPS	eGeMAPS	FAUs	AudioVisual
2 - Class	UL v/s S	97.8	98.6	**99.3**	81.8	99.2
	VL v/s S	95.0	94.3	**98.5**	77.9	98.4
	SL v/s S	82.9	89.1	**93.0**	74.9	88.8
	AL v/s S	**99.5**	97.1	97.1	80.3	97.1
3 - Class	UL/VL/SL	72.6	**74.9**	73.0	50.9	71.5
4 - Class	UL/VL/SL/S	64.9	67.3	**72.5**	50.6	68.2

UL: Unvoiced Laughter, VL: Voiced Laughter, SL: Speech Laughter, AL: All Laughter, S: Speech.

dencies we made use of Leave-One-Speaker-Out (LOSO) cross validation and optimised the training using different solvers and varying the complexity parameter.

4.1 Laughter Recognition

The aim in this set of experiments is to distinguish laughter from speech and distinguish between different kinds of laughter using the audio, visual and audiovisual features. eGeMAPS performed best for the distinction between all types of laughter and speech instances, whereas the MFCCs acoustic feature set performed best for differentiating speech laughter from speech. Even though visual features performed much better than chance on the binary classification tasks, they did not bring any additional improvement in the early fusion, since the acoustic features already provided very high recognition rates. From the results we can conclude that the different categories of laughter we defined can be well identified from speech segments using either audio or video feature set, and differentiating the categories of laughter performs better when using only the audio feature set since the information is mostly conveyed by the auditory channel.

4.2 Emotional Laughter Recognition

In this section we analyze the performance of automatic recognition of emotions in the different categories of laughter we studied. Binary emotion labels (negative/positive) were assigned to each class of laughter/speech based on the mean ratings of arousal and valence calculated using the delayed gold-standard. Results obtained show that voiced laughter performed best for both arousal (77.1%) and valence (81.0%) when using audiovisual features; speech laughter performed equally well on arousal. Performance reported on speech laughter is generally slightly below the one reported for voiced laughter for valence. Whereas results reported for unvoiced laughter show that such episodes convey much less emotion variability, especially for arousal, where only chance level is reported (50.0%). The performance obtained when using all instances of laughter is therefore lower because of the impact of unvoiced laughter. Interestingly, results obtained on speech utterances that do not include any laughter is slightly above the chance level for both arousal (52.8%) and valence (50.9%).

Table 3: Results (%UAR) for 2-class (negative/positive) emotion classification for various categories of laughter; best results over the feature sets are highlighted in bold format.

Category	MFCCs		GeMAPS		eGeMAPS		FAUs		AudioVisual	
	Aro.	Val.	Aro.	Val.	Aro.	Val.	Aro.	Val.	Aro.	Val.
UL	50.0	59.4	50.0	58.2	50.0	**61.0**	50.0	60.7	**50.0**	54.2
VL	49.4	60.7	66.8	63.1	69.6	66.7	47.0	71.1	**77.1**	**81.0**
SL	75.4	63.6	74.1	71.1	**77.1**	69.0	75.7	**74.8**	**77.1**	74.5
AL	50.0	**61.1**	50.7	51.1	50.0	58.0	50.0	59.9	**50.2**	57.2
S	52.8	**50.9**	52.8	50.5	52.8	50.5	52.8	50.5	**52.8**	50.5
All	**51.8**	**50.6**	49.9	50.5	47.9	50.5	**51.8**	**50.6**	51.7	50.5

UL: Unvoiced Laughter, VL: Voiced Laughter, SL: Speech Laughter, AL: All Laughter, S: Speech.

4.3 Context Recognition

Since laughter can be produced in a large variety of contextual situations, we investigated two cases: *spontaneous laughter* v/s *acted laughter* (case 1), and *induced laughter* v/s *spontaneous laughter* (case 2). We selected the MAHNOB (Petridis et al., 2013) database for case 1 since it provide both types of laughter produced by the same subjects. The visual feature set (78.4%) performs better than the acoustic feature sets (75.6%) and the performance increases significantly when both are combined (82.7%). The stereotype of laughter being associated with joy (Duchenne display) could explain the superiority of visual features over audio features, since most of the acted laughter would involve the participants producing laughter with smile.

For Case 2, we perform cross-corpora experiment since there is no database that provides both induced and spontaneous laughter instances. We fused the instances from the six datasets which include interactive data (RECOLA, SEMAINE, BELFAST) and induced data (MAHNOB, AVLC, BINED) and divided them into training and testing data (70:30 ratio). Pure acoustic features

(MFCCs) achieves a very high recognition rate, which might be helped by the acoustic variability present in the different used corpora (microphones, rooms), despite applying a z-score on the features for each dataset. However, the information extracted from the face, which is less subject to cross-corpora variabilities compared to speech, shows that performance is far above the chance level (79.5%).

Table 4: Results (%UAR) for 2 class classification task between Spontaneous and Acted laughter from MAHNOB and for laughter context recognition from speech (2 class) with six datasets; best results over the feature sets are highlighted in bold format.

Case	MFCCs	GeMAPS	eGeMAPS	FAUs	Audiovisual
1	75.6	72.5	72.3	78.4	**82.7**
2	92.6	93.0	**93.9**	79.5	92.2

5 Conclusion

We have provided insights on the automatic analysis of emotional laughter by extensive evaluations carried out on the RECOLA database. Annotations of laughter have been performed on this dataset, and will be made publicly available to the research community. We have then evaluated how the different annotated categories of laughter can be automatically differentiated from audiovisual features, where very high recognition rates have been reported for various acoustic feature sets. Further, we have performed emotion recognition experiments on each of those categories. Results have shown that voiced laughter contains most of the emotion variabilities for both arousal and valence in classification tasks, i. e., passive vs. active, and negative vs. positive. Future work will investigate how variabilities in the language and culture might impact performance on the automatic recognition of laughter.

References

J. Bachorowski, M. Smoski, and M. Owren. 2001. The acoustic features of human laughter. *The Journal of the Acoustical Society of America*, 110(3):1581–1597.

J.-A. Bachorowski and M.J. Owren. 2001. Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science*, 12(3):252–257.

T. Baltrušaitis, M. Mahmoud, and P. Robinson. 2015. Cross-dataset learning and person-specific normalisation for automatic action unit detection. In *Proc. 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 6, pages 1–6. IEEE.

L. Devillers and L. Vidrascu. 2007. Positive and negative emotional states behind the laughs in spontaneous spoken dialogs. In *Proc. Interdisciplinary workshop on the phonetics of laughter*, page 37.

F. Eyben, F. Weninger, F. Gross, and B. Schuller. 2013. Recent developments in openSMILE, the munich open-source multimedia feature extractor. In *Proc. ACM Multimedia (MM)*, pages 835–838. ACM.

G. McKeown, W. Curran, J. Wagner, F. Lingenfelser, and E. André. 2015. The belfast storytelling database: A spontaneous social interaction database with laughter focused annotation. In *Proc. 2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015, Xi'an, China, September 21-24, 2015*, pages 166–172. IEEE Computer Society.

G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schröder. 2012. The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing*, 3(1):5–17.

R. Niewiadomski, M. Mancini, T. Baur, G. Varni, H. Griffin, and M.S.H. Aung. 2013. MMLI: Multimodal multiperson corpus of laughter in interaction. In *Proc. International Workshop on Human Behavior Understanding*, pages 184–195. Springer.

S. Petridis, B. Martínez, and M. Pantic. 2013. The MAHNOB laughter database. *Image and Vision Computing*, 31(2):186–202.

F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne. 2013. Introducing the RECOLA Multimodal Corpus of Remote Collaborative and Affective Interactions. In *Proc. 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE.

I. Sneddon, M. McRorie, G. McKeown, and J. Hanratty. 2012. The Belfast induced natural emotion database. *IEEE Transactions on Affective Computing*, 3(1):32–41.

J. Trouvain. 2003. Segmenting phonetic units in laughter. In *Proc. of the 15th International Conference of Phonetic Sciences*, pages 2793–2796. Barcelona Spain.

J. Urbain, E. Bevacqua, T. Dutoit, A. Moinet, R. Niewiadomski, C. Pelachaud, B. Picart, J. Tilmanne, and J. Wagner. 2010. The AVLaughterCycle Database. In *Proc. LREC*.

Not only decibels: Exploring human judgments of laughter intensity

Magdalena Rychlowska, Gary McKeown,
Ian Sneddon, and William Curran
Queens University, Belfast

Abstract

While laughter intensity is an important characteristic immediately perceivable for the listeners, empirical investigations of this construct are still scarce. Here, we explore the relationship between human judgments of laughter intensity and laughter acoustics. Our results show that intensity is predicted by multiple dimensions, including duration, loudness, pitch variables, and center of gravity. Controlling for loudness confirmed the robustness of these effects and revealed significant relationships between intensity and other features, such as harmonicity and voicing. Together, the findings demonstrate that laughter intensity does not overlap with loudness. They also highlight the necessity of further research on this complex dimension.

1 Introduction

> "Then it came – real laughter, total laughter – sweeping us off in unbounded effusion. Bursts of laughter, laughter rehashed, jostled laughter, laughter defleshed, magnificent laughter, sumptuous and wild… And we laughed to the infinity of the laughter of our laughs… O laughter! Laughter of delight, delight of laughter. Laughing deeply is living deeply." – Milan Kundera

Representations of intense laughter abound in the literature and in cinema and, for most people, such laughs do not need to be defined. We intuitively understand what *unbounded*, *sumptuous*, *wild*, or *deep* laughter is, and we can recognize an intense laugh when we hear it. Research also shows that judgments of laughter intensity are associated with its spontaneous (versus volitional) production (Lavan et al., 2016) and with the extent to which laughs are perceived as humorous (McKeown and Curran, 2015). But what exactly does it mean for laughter to be intense? Empirical investigations of this question have been relatively scarce and methods of measuring intensity are far from consistent. For example, this construct has been described as a function of loudness (e.g. Grammer and Eibl-Eibesfeldt, 1990), emotional arousal (e.g. Urbain et al., 2014), facial movements – measured with facial electromyography (e.g. Hess et al., 1995) or using the Facial Action Coding System (Ekman et al., 1982; Lynch, 2010). Laughter intensity has also been associated with specific patterns of facial expressions, postures, body movements, and physiological changes (Ruch, 1993; Ruch and Ekman, 2001).

In their recent research, McKeown and Curran (2015), adopted an alternative approach and focused on human judgments of laughter intensity. They theorized that, because people often laugh and observe – as well as interpret – laughs produced by others, they are natural "laughter experts". Therefore, these researchers asked a large number of human participants to rate the intensity of laughs presented as audio-visual clips. Here, we build on this previous study and examine the acoustic correlates of human judgments of laughter intensity. Specifically, we analyze ratings of intensity as a function of eleven acoustic dimensions extracted using the PRAAT software (Boersma and Weenink, 2010). We also investigate the predictive power of these dimensions as a function of loudness.

2 Method

2.1 Intensity Ratings

The present research used a subset of data collected by McKeown and Curran (2015). In this study, participants recruited through Amazon's

Mechanical Turk used a scale ranging from 1 to 10 to rate the intensity of laughter sequences, embedded in a website and derived from the Belfast Storytelling Database (McKeown et al., 2015) – a corpus of naturalistic interactions between groups of three or four participants, sitting around a table and telling each other stories. Recorded participants were unaware that laughter was the focus of the research. High-quality audio recordings were collected with head-mounted microphones. Segments of laughter were then extracted from the full recordings of the interaction, removing as much of the contextual information as possible. The present study used intensity ratings of 266 laugh clips of two English-speaking individuals, one male (N = 125) and one female (N = 141), recorded during two separate sessions. Audio sequences were evaluated by 540 Mechanical Turk participants who provided 1490 unique ratings. Each subject rated a variable number of sequences randomly drawn from the entire pool. The number of ratings per clip ranged from 1 to 13 (M = 5.60, SD = 2.52). Ratings of each sequence were then averaged across participants to obtain a single intensity score for each laugh.

2.2 Acoustic Dimensions

The acoustic analysis aimed to cover a broad range of dimensions previously associated with judgments of laughter (e.g. Lavan et al., 2015; Wood et al., 2017). After trimming the silence from the beginning and end of the samples, we used PRAAT (Boersma & Veenink, 2018) to extract eleven acoustic features:

- *Duration* (log-transformed to correct for positive skew),
- L*oudness*, or intensity, in dB,
- Pitch (F0) variables (calculated using the PRAAT autocorrelation algorithm and expressed in semitone scales): *Mean F0* or mean fundamental frequency, *F0 range* (expressed as F0 minimum/F0 maximum), *SD F0/duration*, or the standard deviation of F0 divided by the total duration (log-transformed), and *F0 slope*, or the mean absolute F0 slope (log-transformed). Pitch variables were available for 254 out of 266 laugh sequences.
- Spectral variables: *Center of gravity* (log-transformed), *harmonicity* or harmonics-to-noise-ratio, and *voicing,* or the proportion of voiced frames, versus frames lacking harmonic structure,
- Formant variables: *F1 mean* and *F2 mean*, or the first and second formant.

3 Results

3.1 Correlates of Intensity Judgments

A series of Spearman correlation analyses using raw scores revealed that ratings of laughter intensity were significantly and positively associated with duration, $\rho(264) = .42$, $p < .001$, 99% CI [.28, .56], loudness, $\rho(264) = .43$, $p < .001$, 99% CI [.30, .56], mean F0, $\rho(252) = .41$, $p < .001$, 99% CI [.27, .54], F0 range, $\rho(252) = .41$, $p < .001$, 99% CI [.26, .55], F0 slope, $\rho(251) = .27$, $p < .001$, 99% CI [.12, .42], and center of gravity, $\rho(264) = .21$, $p < .001$, 99% CI [.06, .39]. There was also a significant negative correlation between intensity and F2 mean, $\rho(264) = -.14$, $p = .02$, 99% CI [-.30, .03]. No other correlation was significant, ρs < .10, ps > .10.

Additional analyses revealed a similar correlation pattern for the male and the female expresser, with two exceptions: Center of gravity was a significant predictor of intensity for the male, but not for the female expresser, $\rho(123) = .32$, $p < .001$, 99% CI [.08, .52] and $\rho(139) = .08$, $p = .35$, 99% CI [-.15, .30] The opposite was true for the F2 mean, significantly predicting laughter intensity for the female, but not the male speaker, $\rho(139) = -.26$, $p < .01$, 99% CI [-.45, .04] and $\rho(123) = -.07$, $p = .47$, 99% CI [-.28, .16], respectively.

3.2 Beyond Loudness

The second set of analyses aimed to further explore the relationships described above and test their robustness while controlling for loudness. Our second goal was to test whether the importance of acoustic features as predictors of laughter intensity varies as a function of loudness. In other words, we explored determinants of perceived laughter intensity in soft, versus loud, laughs. For this, we conducted a series of linear regressions, in which participants' ratings of intensity were regressed on interactions between loudness and each of the following acoustic features: log-transformed duration, F0 mean, F0 range, log-transformed SD F0/duration, F0 slope, log-transformed center of gravity, harmonicity, voicing, F1 mean, and F2 mean. Each model used mean-centered variables and included the two main effects along with the interaction.

Table 1 displays the regression statistics. Significant effects are in color: main effects are highlighted in green and interactions – in orange. Asterisks indicate log-transformed variables. Because we estimated 10 unique models, we report p-values adjusted for the false discovery rate.

Variable	**B**	**SE**	**F**	**Adj. p**
Duration*	1.49	.20	56.27	<.001
x Loudness	.09	.02	19.23	<.001
F0 mean	.06	.01	18.08	<.001
x Loudness	<.01	<.01	1.23	.34
F0 range	3.64	.62	33.84	<.001
x Loudness	.14	.06	4.78	.05
SD F0/Duration*	-.52	.27	3.70	.09
x Loudness	-.03	.03	1.28	.34
F0 slope	<.01	<.01	5.19	.04
x Loudness	<.01	<.01	4.11	.07
Center of gravity*	2.43	.38	40.76	<.001
x Loudness	.13	.03	16.60	<.001
Harmonicity	-.11	.02	26.88	<.001
x Loudness	<.01	<.01	0.24	.64
Voicing	-1.43	.39	13.46	<.001
x Loudness	-.02	.03	0.57	.53
F1 mean	<.01	<.01	8.64	<.01
x Loudness	<.01	<.01	0.22	.64
F2 mean	<.01	<.01	0.38	.60
x Loudness	<.01	<.01	2.16	.20

Table 1: Main effects of acoustic features on perceived laughter intensity and their interactions with loudness

The regression analyses revealed that human judgments of laughter intensity were significantly and positively associated with duration, F0 (pitch) mean, range, and slope, with center of gravity and F1 mean. Moreover, we observed a negative relationship of intensity with harmonicity and voicing, such that higher proportions of voiced frames predicted lower ratings of intensity. A closer inspection of the three significant interaction effects showed that the associations between intensity and duration, F0 range, and center of gravity were stronger for loud than for soft laughs. However, each of the three variables as well as F0 mean, F0 slope, harmonicity, voicing, and F1 mean predicted judgements of intensity while controlling for loudness and its interactions.

4 Discussion

The present research investigated acoustic features associated with human judgements of laughter intensity. Our findings reveal that laughs perceived as intense are longer, have higher fundamental frequency, or pitch, higher pitch ranges and steeper pitch slopes (or sharper pitch changes) than laughter rated as low in intensity. This result is consistent with previous research, linking pitch variables with enjoyment and intensity (e.g. Lavan et al., 2016; Niewiadomski et al., 2012; Wood et al., 2017). Intense laughs also have higher centers of gravity, sounding brighter, and are lower in harmonicity, having a less tonal sound than low-intensity laughs (Lavan et al., 2016). They also feature high F1, previously associated with arousal (Laukka et al., 2005). While at the first pass the negative relationship between laugh intensity and the proportion of voiced frames may seem surprising, given the previous evidence that voiced laughs elicit positive affect (Bachorowski and Owren, 2001), it also replicates other, more recent studies linking increased rates of unvoiced segments with spontaneity (Bryant and Aktipis, 2014; Lavan et al., 2015). The difference with the results of Bachorowski and Owren (2001) could also be explained by the fact that we analyzed laughs produced in social interactions, while the corpus used by Bachorowski and Owren features laughs of people watching funny video clips. The use of social laughter also distinguishes our research from the study of Niewiadomski et al. (2012), which investigated acoustic and visual predictors of intensity judgments using laughs produced by participants in reaction to comedy videos.

In sum, our findings reveal that human judgments of laughter intensity are a complex dimension that should not be conflated with loudness. Multiple acoustic features including duration, pitch variables, center of gravity, harmonicity, voicing, and F1 mean predict laughter intensity over and above loudness. These relationships provide insights into why and how relatively soft laughs can be perceived as high in intensity.

One important limitation of the present research is that we only analyzed laughs of two models, which does not allow to draw strong conclusions about the generalizability of the findings.

Describing a larger corpus and including the analysis of laughs produced in different social situations is a work in progress. We also hope to encourage further studies of laughter intensity and the links between this dimension and social motives and feeling states. In the light of recent findings suggesting that, after controlling for intensity, laughs produced in different situations become interchangeable (Curran et al., 2018), examining judgments of laughter intensity has the potential to improve our understanding of social and emotional messages conveyed by this expressive display.

References

Jo-Anne Bachorowski and Michael J. Owren. 2001. Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science*, 12(3): 252-257. doi: 10.1111/1467-9280.00346

Paul Boersma, and David Weenink. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.39, retrieved 20 April 2018 from http://www.praat.org/.

Gregory A. Bryant, and C. Athena Aktipis. 2014. The animal nature of spontaneous human laughter. *Evolution and Human Behavior,* 35: 327-335. doi: 10.1016/j.evolhumbehav.2014.03.003

William Curran, Gary J. McKeown, Magdalena Rychlowska, Elisabeth André, Johannes Wagner, and Florian Lingenfelser. 2018. Social context disambiguates the interpretation of laughter. *Frontiers in Psychology*, 8: 2342. doi: 10.3389/fpsyg.2017.02342

Paul Ekman, Wallace V. Friesen, and Joseph C. Hager. 2002. *Facial Action Coding System: The manual on CD-ROM.* Salt Lake City, UT: Research Nexus.

Karl Grammer and Irenaus Eibl-Eibesfeldt. 1990. The ritualization of laughter. In *Natürlichkeit der Sprache und der Kultur: acta colloquii*, pages 192-214, Brockmeyer, Bochum, Germany.

Ursula Hess, Rainer Banse, and Arvid Kappas. 1995. The intensity of facial expression is determined by underlying affective state and social situation. *Journal of Personality and Social Psychology*, 69(2): 280-288. doi: 10.1037/0022-3514.69.2.280

Milan Kundera. 1986. The book of laughter and forgetting. Penguin Books, New York, NY.

Petri Laukka, Patrik Juslin, and Roberto Bresin. 2005. A dimensional approach to vocal expression of emotion. *Cognition and emotion*, 19(5): 633-653. doi: 10.1080/02699930441000445

Nadine Lavan, Sophie Scott, and Carolyn McGettigan. 2015. Laugh like you mean it: Authenticity modulates acoustic, physiological and perceptual properties of laughter. *Journal of Nonverbal Behavior*, 40(2): 133-149. doi: 10.1007/s10919-015-0222-8

Robert Lynch. 2010. It's funny because we think it's true: Laughter is augmented by implicit preferences. *Evolution & Human Behavior,* 31(2): 141–148. doi: 10.1016/j.evolhumbehav.2009.07.003

Gary J. McKeown and William Curran. 2015. The relationship between laughter intensity and perceived humor. In *Proceedings of the 4th Interdisciplinary Workshop on Laughter and Other Non-Verbal Vocalisations in Speech*, pages 27-29.

Gary McKeown, William Curran, Johannes Wagner, Florian Lingenfelser, and Elisabeth André. 2015. The Belfast Storytelling Database: A spontaneous social interaction database with laughter focused annotation. Paper presented at 6th International Conference on Affective Computing and Intelligent Interaction, Xi'an, China. doi: 10.1109/acii.2015.7344567

Radoslaw Niewiadomski, Jérôme Urbain, Catherine Pelachaud, and Thierry Dutoit. 2012. Finding out the audio and visual features that influence the perception of laughter intensity and differ in inhalation and inhalation phases. In *Proceedings of the 4th International Workshop on Corpora for Research on Emotion, Sentiment, and Social Signals, Istanbul, Turkey.*

Willibald Ruch. 1993. Exhilaration and humor. In *Handbook of emotions*, vol. 1, pages 605–616. Guilford Press, New York, NY.

Willibald Ruch and Paul Ekman. 2001. The expressive pattern of laughter. In *Emotion, qualia, and consciousness*, pages 426-443. World Scientific, Tokyo, Japan.

Jérôme Urbain, Hüseyin Çakmak, Aurélie Charlier, Maxime Denti, Thierry Dutoit, and Stéphane Dupont. 2014. Arousal-driven synthesis of laughter. *IEEE Journal of Selected Topics in Signal Processing,* 8(2): 273-284. doi: 10.1109/JSTSP.2014.2309435

Adrienne Wood, Jared Martin, and Paula Niedenthal. 2017. Towards a social functional account of laughter. Acoustic features convey reward, affiliation, and dominance. *PLoS ONE,* 12(8): e0183811. doi: 10.1371/journal.pone.0183811.

On Laughter Intensity Level: Analysis and Estimation

Kevin El Haddad, Huseyin Cakmak, Thierry Dutoit
numediart Institute, University of Mons /31 Boulevard Dolez, Mons, Belgium
{kevin.elhaddad, huseyin.cakmak, thierry.dutoit}@umons.ac.be

Abstract

This work focuses on laughter intensity level, the way it is perceived and suggests ways to estimate it automatically. In the first part of this paper, we present a laughter intensity database which is collected through online perception tests. Participants are asked to rate the overall intensity of laughs. Presented laughs are either audio only or visual only or audiovisual. Statistical analysis show that the perceived intensity is significantly higher when the modality is visual only and suggests that the audio cue might have the biggest influence on laughter intensity perception.We also show that the order by which the modalities are presented to the raters may influence the perception of laughter intensity. In the second part, different estimation/classification techniques were tested including GMM-based mapping and common classification techniques. A set of features were defined, extracted and tested for classification. Results show that the estimation of the global audio laughter intensity is possible with good classification performances.

1 Introduction

Laughter is everywhere. So much that we often do not even notice it. It is in common believes and true that laughter has a strong connection with humor (G. and W., 2015). Most of us seek out laughter and people who make us laugh, and use it in our gatherings and social interactions. Laughter also plays an important role in making sure we interact with each other smoothly. It provides social bonding signals that allow our conversations to flow seamlessly between topics; to help us repair conversations that are breaking down; and to end our conversations on a positive note. In the last decades, with the development of human-machine interactions and various progress in speech processing, laughter became a signal which machines should be able to detect, analyze and produce. This work focuses on the estimation of laughter intensity from acoustic features.

In 2001, Ruch and Ekman (Ruch and Ekman, 2001) published an extensive report on the production of laughter. They investigated various aspects like phonation, respiration, muscular and facial activities. Laughter is described as an inarticulate utterance. Its cycle is around 200 ms and it is usually operated on the expiratory reserve volume. The same year, Bachorowski et al. (Bachorowski et al., 2001) focused on the acoustic properties of human laughter and its differences with speech. They found that laughter yields higher fundamental frequencies than speech, formant frequencies in laughter correspond to central vowels and unvoiced laughter accounts for 40 to 50% of laughter occurrences. Chafe (Chafe, 2007) also describes the mechanical production of laughter and presents various acoustic laughter patterns. A common conclusion of these studies is the high variability of the laughter phenomenon, in terms of voicing, fundamental frequency, intensity and, more generally, types of sounds (grunts, cackles, pants, snort-like sounds, etc.).

Intensity is an important dimension of laughter. The notion of intensity seems so natural that most researchers do not define it (e.g., (Glenn, 2003; Chafe, 2007; Edmonson, 1987)). In (Ruch, 1993), Ruch defines the emotion of exhilaration, which is one of the emotions leading to laughter. He discusses different levels of intensity of this emotion and the corresponding behaviors, from smile at low intensity to laughter accompanied by posture changes (throwing back the head, vibrations of the trunk and shoulders) at high intensity. Furthermore, intensity is encoded differently by individuals, with reference to their own laughing style (Edmonson, 1987).

Since intensity is a fundamental dimension, frequently and naturally used to describe laughs, it appears as an important feature to be able to estimate for further use in laughter synthesis (Urbain et al., 2014) or recognition. Indeed, it can give valuable information about the state of a participant in a human-machine interaction system. It is also a convenient layer in interactive systems to separate the processes of deciding to laugh (with a target intensity), which is independent from the laughter synthesis voice and style, and synthesizing the corresponding laugh, which obviously depends on the modeled individual traits. In this paper we will use the term *intensity* to refer to the intensity level of the laughter perceived by a listener.

This paper revolves around laughter intensity, how it is perceived and proposes a machine learning based method to detect it.

This paper is organized as follows : Section 2 gives details on the intensity data collection process, Section 3 provides an analysis of the collected data, Section 4 presents a method for laughter intensity level estimation and the experiments leading to it. Finally, Section 5 concludes the paper and give future work perspectives.

2 Online perception tests

To collect the intensity data, online tests were conducted. Participants were asked to rate the intensity of laughs on a 5-point scale ranging from 0 to 4. Laughs from 3 subjects were evaluated in this test. Two subjects (1 male, 1 female) from the AVLC Database (Urbain et al., 2010) and one subject (male) from the AVLASYN Database (Çakmak et al., 2014). A total of 334 laughs were used. The number of laughs from each of the subjects is given in Table 1.

Table 1: Number of laughs for each subject in the experiment

AVLC DB (Subject 6)	67
AVLC DB (Subject 14)	65
AVLASYN DB (D4)	202
TOTAL	334

Due to relevant data availability, the laughs from the AVLC Database were evaluated only on the audio while laughs from the AVLASYN database were evaluated along 3 different modalities; audio only, video only (video without sound) and both together.

In the case of the AVLASYN Database, 7331 ratings were collected from 226 participants (135 males and 91 females from 18 to 77 years old with an average age of 31.46 and a standard deviation of 10.23).

Table 2 gives the average number of time each file has been evaluated in each of the 3 parts.

Table 2: Average number of time each file has been evaluated in each part of the test

Modality	Average (std)
Audio only	11.78 (3.47)
Video only	12.03 (3.30)
Audiovisual	11.95 (3.48)

In the case of the AVLC Database, 1505 evaluations were collected from 40 participants (32 males and 8 females from 20 to 61 years old with an average age of 35.38 and a standard deviation of 10.43). The pipeline followed for the test is the same as above but it contains only one part which is the audio only and each participant was asked to evaluate 40 laughs. Each file has been evaluated 11.40 times on average with a standard deviation of 2.96.

3 Data Analysis

3.1 Analysis of the perceived intensity in each specific modality

Our first experiment focuses on the possible difference between the perceived intensity with respect to the different modalities in the case of the AVLASYN Database. If we calculate the Pearson's correlation coefficients on the mean intensity values obtained for each single laugh in the different modalities that have been tested, we obtain the following matrix :

$$\left[\begin{array}{cccc} & Audio & Video & AV \\ Audio & 1.0000 & 0.9002 & 0.9654 \\ Video & 0.9002 & 1.0000 & 0.9185 \\ AV & 0.9654 & 0.9185 & 1.0000 \end{array}\right]$$

As expected, there is a strong correlation between the cases. However, we see that the correlation is even stronger between Audio only and Audivisual than between Audio only and Video only. The mean and standard errors of intensity scores of each part are given in Table 3.

Table 3: Mean and standard errors of intensity scores for each part

Part	Mean (std. err.)
Audio only	1.80 (0.0049)
Video only	2.00 (0.0044)
Audiovisual	1.80 (0.0048)

This suggests that there might be a difference in the perception of laughter intensity when audio is not present.

3.1.1 Analysis of variance on modalities

To verify this hypothesis, we have conducted an ANOVA test with a post-hoc TUKEY Honest Significant Difference analysis with a confidence level of 99% between the results to the different parts (modalities) of the online test. The pairwise p-values are given below with significant differences in bold :

$$\left[\begin{array}{cccc} & Audio & Video & AV \\ Audio & - & \mathbf{0.00} & 0.18 \\ Video & \mathbf{0.00} & - & \mathbf{0.00} \\ AV & 0.18 & \mathbf{0.00} & - \end{array}\right]$$

The p-values comparison shows that there is a significant difference between the visual only modality and the two others. This confirms our thoughts that the visual modality alone is perceived differently than the case with audio. The mean scores suggest that the visual modality alone tends to be perceived with a higher intensity. Of course, these conclusions are valid only for the studied subject and it might be interesting to investigate the possible generalization of these findings to any laughs or to specific categories of laugh.

3.1.2 Analysis of variance on tests order

One other analysis which may also be interesting is the possible influence of the order in which the modalities are presented to a given participant. As explained above, the perception test is such that 3 different types (audio only, visual only and audiovisual) of files were presented in 3 different successive parts of the test and the order was randomly determined when the test begins. To assess whether or not the order in which the different parts are presented has an influence on the perceived intensity, we perform a One-way ANOVA and the TUKEY HSD post-hoc analysis. P-values are given in Table 4. In this table, for the ease of read, the sequence are defined by 3 numbers. Each number referring to a specific modality ; 1 is for the audio only test, 2 is for the visual only test and 3 is for the audiovisual test. A sequence referred as 123 therefore means that the underlying order of the test was audio only then visual only and finally audiovisual. The main conclusion from this table is that there is a statistically significant influence of the order of the tests on the perceived intensity. It is however hard to find clear patterns from specific test sequences. It is reasonable to think that the position of the video only modality in the test order may have an influence. Indeed, in that modality, the intensity is perceived differently as shown in the previous section.

3.2 Analysis of the intensity of each studied subject

Figure 1 gives the boxplots of the intensity values for each subject. We can see that the Subject 14 (female) has the median, 25th percentile and 75th percentile clearly lower than the two other male subjects. The two male subject has similar medians (2.0 and 1.9) and 25th percentiles (both 1.0). However, the 75th percentile is higher for Subject 6 (3.06 against 2.64). Maximum and minimum values are similar for all subject with Subject 6 slightly higher though.

4 Audio laughter intensity estimation

Among the possible applications of the intensity information presented in this paper, there is the

Table 4: Pairwise comparison p-values for the different orders in which the test were presented. Significant differences with a confidence level of 95% are given in bold.

Compared Test Order Pairs	p-values
132-123	0.37
213-123	0.20
231-123	0.77
312-123	1.00
321-123	0.17
213-132	**0.00**
231-132	**0.01**
312-132	0.63
321-132	**0.00**
231-213	0.91
312-213	**0.04**
321-213	1.00
312-231	0.38
321-231	0.89
321-312	**0.03**

estimation of the intensity of a given audio laughter file. It is also important to note that, as shown in this paper, there is not a statistically significant difference between the perceived intensity of an audio only laughter and the same audiovisual laughter. Therefore, we can estimate the intensity of a given audiovisual laugh based on the acoustic information only.

To do this, we propose here to use a Gaussian Mixture Model (GMM) based approach. First, silences are removed from the input audio laughter files. Then, a set of features are extracted from these files and the features that are the most correlated with the output intensity levels are kept. The selected features are then used to train GMMs with full covariance matrices. Doing so, we can model the relationship between the input acoustic features and the corresponding intensity levels. The GMM mapping framework used in this work was first introduced in 1996 by Stylianou (Stylianou, 1996) for voice conversion. The implementation used here is the one of Kain (Kain, 2001) also used in recent work such as (Hueber et al., 2011).

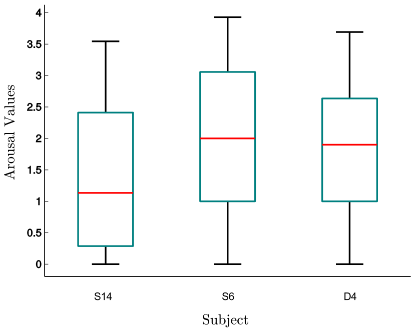


Figure 1: Boxplots for each studied subject. The median is given in red inside the boxes, 25th and 75th percentiles are the limits of the main boxes and the upper and lower tails give the minimum and maximum values of the distribution.

4.1 Feature selection

A set of features are extracted from the audio files. Some features are scalar values related to the whole file in the first place while others are continuous features extracted using 10ms windows and 25ms frame shift. The list of extracted features are as follows :

- Spectrogram
- Acoustic Features listed in Table 5 from (Giannakopoulos and Pikrakis, 2014)
- Fundamental Frenquency (F0) extracted using Straight (Kawahara, 2006)

Table 5: List of the 36 features from (Giannakopoulos and Pikrakis, 2014)

- Zero Crossing Rate	1 dim
- Energy	1 dim
- Energy Entropy	1 dim
- Spectral Centroid	2 dim
- Spectral Entropy	1 dim
- Spectral flux	1 dim
- Spectral Rolloff	1 dim
- MFCCs	13 dim
- Harmonic Features	2 dim
- Chroma Vector	12 dim
- Spectral Zone	1 dim

Since all these features are continuous features, we derived the following descriptors related to the whole file : mean, standard deviation, range, root mean square and histogram values of each feature ; the mentioned histogram values are the number of elements in each of the bins of a histogram calculated on each continuous feature by imposing the number of bins to 3. Among the most correlated features, we mainly find F0 related features, Chroma vector related features, the mean of the zero-crossing rate and energy entropy standard deviation.

4.2 Results

We define 4 different cases of training and testing sets as detailed in Table 6. Case 1 is a training on all the data following a leave-one-out approach. Cases 2 and 3 are used to assess the performances when testing on a subject that was not seen at all in the training. Case 4 is to try if performances are improved when a few examples of the testing subject are shown in training. In this table, the available 3 subjects are named S6 and S14 for subjects 6 and 14 from the AVLC Database and D4 for the subject from AVLASYN Database. Table 7 gives the accuracy results for each case. The accuracy is defined as the number of files for which the intensity estimation error is less than 0.5 (on a scale going from 0 to 4). The table gives all the accuracy values for the cases and sub-cases enumerated in Table 6. We can see that all the accuracy results are over 90% except when the testing is done on the subject S14 (female) for the cases 2 and 3 which correspond to a training on male subjects. We also see that the accuracy increases when we add a few examples of the test subject, even more fore the female subject (see case 4 results).

Table 6: List of train/test sets

	TRAIN SET	TEST SET
1	- D4+S6+S14	- leave-one-out
2	- D4	- S6 - S14
3	- D4+S6 - D4+S14	- S14 - S6
4	- D4+S6+10 files from S14 - D4+S14+10 files from S6	- Remaining of S14 - Remaining of S6

Table 7: Estimation results for each case

CASE	1	2	3	4
ACC. (S6)	96.40%	94.03%	91.04%	92.98%
ACC. (S14)		81.54%	86.15%	90.91%

Table 8: CASE 1 : Best classification results with $\epsilon = 0.5$ (first row) and $\epsilon = 0$ (second row)

GMM	TREE	DIS	KNN	NN	SVM
96.4% (46)	86.3% (12)	91.6% (44)	77.2% (51)	88.6% (81)	**92.9%** (43)
83.8% (41)	39.5% (10)	48.5% (32)	39.8% (51)	54.5% (90)	48.0% (51)

4.3 GMM vs other machine learning methods

In this section we show the reason why the GMM method was chosen by presenting the results of our comparison with other methods. The same training and testing settings as in the previous sections were used here to the binary classification decision tree (TREE), discriminant analysis (DIS), K-Nearest Neighbor (KNN), single layer neural network with 9 neurons (with softmax activation functions and trained with gradient descent) and Support Vector Machine (SVM). To evaluate the classification, we consider that a file is correctly classified with a tolerance (ϵ) of 0.5. This means that if the classification error is at most 0.5 (e.g. 2.5 instead of 2) it is considered as a correct classification. Test were also made with no tolerance in CASE 1 for the sake of comparison. Table 8 shows that GMM has a clear advantage in this respect.

The results for CASE 1 are given in Figure 2 and Table 8. We can see that the best method is clearly the GMM mapping followed by SVM and Discriminant Analysis.

Figure 2: Results for CASE 1 when using the first n most correlated features for training ($n \in [1 : 100]$) and tolerance 0.5

5 CONCLUSION AND FUTURE WORKS

In this paper, we studied the intensity level estimation of an audio laughter file from acoustic features. Results show that the estimation is possible. Among the compared methods, GMM-based mapping appears to be the best in all the tested cases. This method also offers good perspectives on the estimation of intensity values not limited to a finite number of classes. Indeed, GMMs can be used for mapping on decimal values.

In future work we intend to collect to bigger a database of annotated data in order to leverage the power of deep learning. We also intend to link this laughter intensity estimation system with other task such as laughter detection, laughter type classification. This latter will contribute to improve context understanding in intelligent systems.

6 Conclusion

This work was focused on laughter intensity. We tried to understand more about its perception by mainly studying the effect of the modality in the perception process. For this we collected a database of laughs, annotated the intensity level of each laugh via online perception tests and analysed them. The results suggested that the audio cue might have the biggest influence on the perception. But further studies are required to confirm it. We compared several machine learning based systems to estimate laughter intensity and showed that the GMMs outperformed the other methods considered. In the future, we intend to increase our database which would allow the use of more advanced techniques such as recurrent and convolutional neural networks.

References

J.-A. Bachorowski, M. J. Smoski, and M. J. Owren. 2001. The acoustic features of human laughter. *Journal of the Acoustical Society of America*, 110(3):1581–1597.

H. Çakmak, J. Urbain, and T. Dutoit. 2014. The AV-LASYN database : A synchronous corpus of audio and 3d facial marker data for audio-visual laughter synthesis. In *Proc. of the 9th Int. Conf. on Language Resources and Evaluation (LREC'14)*.

Wallace Chafe. 2007. *The Importance of not being earnest. The feeling behind laughter and humor.*, paperback 2009 edition, volume 3 of *Consciousness & Emotion Book Series*. John Benjamins Publishing Company, Amsterdam, The Nederlands.

Munro S Edmonson. 1987. Notes on laughter. *Anthropological linguistics*, pages 23–34.

McKeown G. and Curran W. 2015. The relationship between laughter intensity and perceived humour. In *The 4th international workshop on laughter and other non-verbal vocalizations in speech.*

Theodoros Giannakopoulos and Aggelos Pikrakis. 2014. *Introduction to Audio Analysis: A MATLAB® Approach*. Academic Press.

Phillip J Glenn. 2003. *Laughter in interaction*. Cambridge University Press, Cambridge.

Thomas Hueber, Elie-Laurent Benaroya, Bruce Denby, and Gérard Chollet. 2011. Statistical mapping between articulatory and acoustic data for an ultrasound-based silent speech interface. In *INTERSPEECH*, pages 593–596.

Alexander Blouke Kain. 2001. *High resolution voice transformation*. Ph.D. thesis, Oregon Health & Science University.

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

W. Ruch. 1993. Exhilaration and humor. *Handbook of emotions*, 1:605–616.

W. Ruch and P. Ekman. 2001. The expressive pattern of laughter. In A. Kaszniak, editor, *Emotion, qualia and consciousness*, pages 426–443. World Scientific Publishers.

Ioannis Stylianou. 1996. *Harmonic plus noise models for speech, combined with statistical methods, for speech and speaker modification*. Ph.D. thesis, Ecole Nationale Supérieure des Télécommunications.

J. Urbain, E. Bevacqua, T. Dutoit, A. Moinet, R. Niewiadomski, C. Pelachaud, B. Picart, J. Tilmanne, and J. Wagner. 2010. The avlaughtercycle database. In *Proc. of the Seventh Int. Conf. on Language Resources and Evaluation (LREC'10).*

Jérôme Urbain, Hüseyin Çakmak, Aurélie Charlier, Maxime Denti, Thierry Dutoit, and Samuel Dupont. 2014. Arousal-driven synthesis of laughter. *Selected Topics in Signal Processing, IEEE Journal of*, 8(2):273–284.

Phonetic entrainment of laughter in dating conversations: On the effects of perceived attractiveness and conversational quality.

Jan Michalsky & Heike Schoormann

Institute of German Studies, University of Oldenburg, Germany
j.michalsky@uol.de, heike.schoormann@uol.de

Abstract

Laughter serves as a signaling device to facilitate the establishment and maintenance of human relationships. The role of laughter in mating conversations, however, seems to be as crucial as it is still underinvestigated. In this study we seek to examine how a speaker's perception of an interlocutor in terms of visual attractiveness as well as his/her perception of the quality of the conversation affect the phonetics of laughter in dating conversations. In addition to the phonetics of laughter in absolute terms, we study laughter entrainment, i.e. the mutual influence of laughing behavior, since previous studies found both attractiveness and conversational quality to affect speech in terms of prosodic entrainment. As this study constitutes work in progress, preliminary results and working hypothesis are reported.

1 Introduction

The investigation of paralinguistic phenomena of the speech signal and their dependence on not only emotional states and attitudes but also conversational settings has gained increasing attention in the past decade. Amongst the many conversational settings, the specific type of mating conversations seems of particular interest since its more basic nature regarding human behavior suggests a stronger connection to the more primal and universal parts of speech found in the paralinguistic domain. In this paper we focus on how the relationship of two speakers in terms of their perception of the respective interlocutor's visual attractiveness as well as their attitude towards the conversation as a whole (henceforth referred to as conversational quality) shapes their paralinguistic behavior and its affective impact in a spontaneous dating conversation.

In previous studies, both perceived visual attractiveness as well as perceived conversational quality have been found to affect the paralinguistic domain. Speakers who perceived their interlocutor as more or less attractive were found to either raise or lower their overall mean fundamental frequency. This effect was speaker sex-specific with male speakers lowering their mean f0 when talking to a more attractive female interlocutor (Hughes, Farley & Rhodes 2010) and female speakers raising their mean f0 (Fraccaro et al. 2011). This resulted in female and male speakers becoming more dissimilar in terms of f0 when talking to a more attractive interlocutor, a phenomenon which is also known as disentrainment (cf. Michalsky & Schoormann 2018). Likewise, conversational quality affects a speaker’s fundamental frequency. However, when perceiving a conversation as more positive or pleasant, speakers were found to adjust their f0 mean to that of their interlocutor by becoming more similar in absolute terms which is known as entrainment (cf. Lee et al. 2010, Michalsky, Schoormann & Niebuhr 2018).

Both findings have been linked to an evolutionary biological explanation. With respect to Ohala’s (1983, 1984) frequency code, the effects of attractiveness can be regarded as a means to signal *largeness* or *smallness* and thus evoke the impression of dominance or attraction. The effects of conversational quality, however, can be associated with a more general connection between linguistic distance and social distance. A higher phonetic similarity resembles closeness between the interlocutors which is in turn related to the perceived conversational quality.

Although the past decade has furthered our understanding of the phonetics and social functions of laughter, most aspects of this nonverbal behavior are still largely unknown. In general, laughter is regarded as a very old and primal form of human vocalization serving the purpose of establishing and maintaining human relationships (cf. Bachorowski 1999, Bacharowski & Owren 2001, Szameitat et al. 2012). Accordingly, it can be assumed that laughter is affected by perceived attractiveness or conversational quality in dating settings similarly to speech (cf. Grammer 1990, Grammer & Eibl-Eibesfeldt 1990, Bachorowski & Orwen 2001, Bachorowski, Smoski & Owren 2001, Szameitat et al. 2012).

Earlier findings by Grammer and Eibl-Eibesfeldt (1990) suggest that laughter greatly differs between same-sex and opposite sex dialogues. Their findings suggest that male speakers and female speakers behave differently with respect to the degree of periodicity found in laughter when engaging in a dating conversation. However, Devillers and Vidrescu (2007) found in their study on different types of laughter that the degree of periodicity varies with whether the laughter is uttered in a positive or negative way. This is in accordance with Bachorowski, Smoski and Owren's (2001) investigation of laughter types which also shows a link between periodicity and positivity in laughter. Secondly, Kipper & Todt (2007) that when uttered without in isolation without an active interlocutor female and male speakers showed no speaker sex-specific differences in the phonetic realization of laughter including f0. Accordingly, the differences in periodicity found by Grammer and Eibl-Eibesfeldt (1990) might not only depend on the mating setting per se but can be traced back to social variables connected to the interlocutor or the conversation, namely perceived attractiveness and conversational quality.

Lastly, Trouvain and Truong (2012), and Truong and Trouvain (2012a, 2012b) found that laughter shows effects of entrainment or convergence comparable to prosodic features mentioned above. (ibd.) showed that speakers do not only change how they laugh in absolute terms but also adjust their laughter with respect to the laughing behavior of their interlocutor with respect to e.g. duration, temporal alignment, and potentially even phonetic characteristics. This might also affect the degree of periodicity as well. Since the degree of f0 entrainment in dating conversations was found to depend on both perceived visual attractiveness and perceived conversational quality, we want to investigate whether this holds for the entrainment of laughter as well.

We arrive at the following research questions: 1) Does the degree of perceived visual attractiveness and/or perceived conversational quality affect the phonetics of laughter in absolute terms in dating conversations? 2) Does the degree of perceived visual attractiveness and/or perceived conversational quality affect the degree to which speakers entrain the phonetics of their laughter to their interlocutor in dating conversations?

2 Method

2.1 Subjects

10 female and 10 mal paid volunteers from the University of Oldenburg participated in the study. All subjects were aged between 19 and 28 years, monolingual speakers of High German and grew up in Lower Saxony. For this study only heterosexual singles were included. All subjects were unacquainted and interaction prior to the experiment conversations was avoided.

2.2 Procedure

All subjects were informed about the dating setting prior to the experiment. Each participant was paired with each participant of the opposite sex for a total of 100 opposite-sex pairs. The subjects were seated in a quiet room and participated in short natural spontaneous conversations of 15 to 20 minutes each with no topic restrictions. All participants judged the visual attractiveness as well as the general likability of their interlocutor immediately before and after each conversation and the general impression they had of the conversation itself only after the conversation on a 10-point Likert scale. These ratings were not revealed to the respective interlocutors. Recordings were made in stereo using a portable digital recorder (Tascam HD P2) at a sampling rate of 48 kHz and 24-bit resolution with head-mounted microphones (DPA 4065 FR).

2.3 Acoustic analysis

In this paper we focus solely on the so called phenomenon of *free laughter* (Kohler 2007) including the more typical segmental realizations as well as vocalized inhalations, glottal pulses, and nasal bouts but excluding both speech laughter and speech smiles. The audio tracks were separated by speaker and for each speaker and each conversation all instances of free laughter were manually annotated and extracted using Praat (Boersma & Weenink 2016). For each conversation the number of laughs was assessed and the total duration of laughter measured in seconds. Lastly, the percentage of locally unvoiced frames was automatically calculated using the Praat voice report function. For the work in progress reported in this paper only 10 % of the data set have been annotated and are used as a sample for the preliminary results.

3 Results

As this paper constitutes work in progress, preliminary results will be presented impressionistically.

3.1 Laughter count and duration

Both the number of laughs and the total duration of laughter within the conversations are strongly correlated with both perceived attractiveness and perceived conversational quality. Although this is not surprising, the degree of these differences is noticeable. We found one speaker showing only five instances of laughter with a total duration of 4.5 s in a conversation which she perceived as unpleasant with an interlocutor she perceived as unattractive and 19 instances with a total duration of 36.8 s in a conversation she perceived as pleasant with an interlocutor she perceived as attractive while both conversations lasted about 15 minutes. This effect is consistent in the preliminary sample.

Furthermore, there is an asymmetry in both number and duration of laughter depending on the individual perception of the conversation and the interlocutor. While perceiving their interlocutor as attractive or the conversation as pleasant increased the number and duration of laughs a speaker uttered, the number of laughs uttered by the interlocutor shows no effect. Accordingly, in situations where there was a crucial gap between the perceptions of two speakers regarding the quality of the conversation and/or the perception of their interlocutor's attractiveness there is also a large difference in the number and duration of laughs.

3.2 Laughter periodicity, attractiveness and conversational quality

According to our research hypothesis we find a high correlation between the fraction of locally unvoiced frames, perceived attractiveness, and conversational quality. Speakers show a lower percentage of unvoiced frames when talking to an interlocutor they perceived as more visually attractive or when talking in a conversation they perceived as more positive.

Furthermore, we find consistent effects for speaker sex with both male and female speakers increasing the number of voiced frames with increasing perceived attractiveness and conversational quality.

Lastly, we again find a crucial asymmetry within the conversations. While conversations with mutually high perceived attractiveness and mutually high perceived conversational quality are characterized by both speakers showing a high degree of periodicity within their laughs, conversations with high differences in the perceived social variables also show large differences in the phonetic properties of the laughs of the two speakers in terms of periodicity.

3.3 Laughter periodicity and entrainment

Although the correlation between periodicity, perceived attractiveness, and conversational quality is fairly consistent even within our small sample, there are certain instances that cannot be explained in this way. There is one example of a conversation where a male speaker showed almost exclusive unvoiced laughter although talking to an interlocutor he evaluated as nine out of ten in terms of perceived attractiveness within a conversation perceived as ten out of ten in terms of quality. However, it is noticeable that the female interlocutor, who did not share his judgements in terms of perceived attractiveness, had a very distinct unvoiced laughter.

4 Discussion

The preliminary results suggest that the number, the duration, and the phonetics of laughter in spontaneous dating conversations are significantly affected by social variables such as the perceived attractiveness of the interlocutor and the perceived quality of the conversation.

The findings regarding the number and duration of laughter show that the effects of social variables on the presence of laughter itself are crucial. In bad conversations with unlikable or unattractive interlocutors, laughter could be almost absent. This becomes even more noteworthy when regarding the asymmetry. Our results suggest that it is perfectly possible that one interlocutor perceives the conversation as bad and shows close to no instances of laughter while this goes completely unnoticed by his/her interlocutor who perceives the conversation as positive and shows a high amount of laughter in contrast. This finding of laughing alone seems to be in sharp contrast to previous findings on joint laughter Trouvain and Truong (2012) but can be explained well through the low degree of conversational quality.

Furthermore, according to our research hypothesis we found a higher degree of periodicity in instances of laughter uttered in conversations that were perceived as positive or with interlocutors that were perceived as more attractive. Although this has to be explored in more depth, this is in accordance with the findings by (Bachorowski, Smoski and Owren (2001) and Devillers and Vidrescu (2007) that laughter in more positive situations is characterized by a higher degree of periodicity and thus the laughter in better conversations with more attractive people can be interpreted as either more positive or more genuine. However, the findings are slightly at odds with the findings by Grammer and Eibl-Eibesfeldt (1990) since in our study both female and male speakers increased the degree of periodicity in the same direction regarding the social variables. Then again, (ibd.) investigated differences between same-sex and opposite-sex conversations and not within different qualities of opposite-sex conversations, which might make the results complementary and not contradicting.

Since there was a strong correlation between perceived attractiveness as well as perceived conversational quality in the sample we chose for this preliminary study, we are not able to tell which of the two qualifies constitutes the decisive factor at this point. This will be a focus of future research.

Lastly, we found instances in our sample were the degree of perceived attractiveness and perceived conversational quality could not explain the high degree of unvoiced frames in the laughter. However, as we pointed out, the potential variable affecting the laughter of the speaker in question could have been the distinctive phonetic quality of the laughter of his interlocutor. A post-hoc investigation of our sample shows that, although asymmetries are found, the majority of the conversations with high judgements for both social variables were mutual and thus showed similar phonetic properties in the instances of laughter. Accordingly, we hypothesize that at least some if not all cases of varying degrees of periodicity in laughter depending on perceived attractiveness and perceived conversational quality might not be explained through an immediate effect of the social variables. Instead, we might observe (an effect of the degree of) phonetic entrainment of laughter between the interlocutors comparable to the effects found for f0 mean as a paralinguistic feature (cf. Szameitat et al. 2012, Trouvain & Truong 2012, Truong & Trouvain 2012a, b). However, this is an assumption that needs to be thoroughly examined by extending the sample to the whole corpus in future research.

References

J.-A. Bachorowski. 1999. Vocal expression and perception of emotion. In *Current Direction in Psychological Sciences*, 8: 53-57.

J.-A. Bachorowski, M. J. Smoski, and M. J. Owren. 2001. The acoustic features of human laughter. In *Journal of the Acoustical Society of America*, 110: 1581-1597.

J.-A. Bachorowski, and M. J. Owren. 2001. Not all laughs are alike; Voiced but not unvoiced laughter elicits positive affect in listeners. *Psychological Science,* 12: 252-257.

P. Boersma, and D. Weenink. 2016. *Praat: Doing phonetics by computer*.

N. Campbell. 2007. Whom we laugh with affects how we laugh. In *Workshop on the Phonetics of Laughter*, pages 61–65.

L. Devillers, and L. Vidrescu, 2007. Positive and Negative emotional states behind the laughs in spontaneous spoken dialogs. In *Workshop on the Phonetics of Laughter*, pages 37–40.

P. J. Fraccaro, B. C. Jones, J. Vukovic, F. G. Smith, C. D. Watkins, D. R. Feinberg, A. C. Little and L. M. Debruine. 2011. Experimental evidence that women speak in higher voice pitch to men they find attractive. In *Journal of Evolutionary Psychology*, 9(1): 57–67.

K. Grammer. 1990. Strangers meet: Laughter and nonverbal signs of interest in opposite-sex encounters. In *Journal of Nonverbal Behavior*, 14(4): 209–236.

K. Grammer and I. Eibl-Eibesfeldt. 1990. The ritualisation of laughter. In W. A. Koch. Die Natürlichkeit der Sprache und der Kultur. Bochum: Brockmeyer, pages 192–214.

S. M. Hughes, S. D. Farley and B. C. Rhodes. 2010. Vocal and physiological changes in response to the physical attractiveness of conversational partners. In *Journal of Nonverbal Behavior*, 34: 1–13.

K. Kohler. 2007. 'Speech-smile', 'speech-laugh', 'laughter' and their sequencing in dialogic interaction. In *Workshop on the Phonetics of Laughter*, pages 21–26.

S. Kipper, and D. Todt. 2007. Series of similar vocal elements as a crucial acoustic structure in human laughter. In *Workshop on the Phonetics of Laughter*, pages 3-8.

C. C. Lee, M. P. Black, A. Katsamanis, A. C. Lammert, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. S. Narayanan. 2010. Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. In *Proceedings of Interspeech*, pages 793–796.

J. Michalsky, and H. Schoormann. 2018a. Opposites attract! Pitch divergence at turn breaks as cause and effect of perceived attractiveness. In *Proceedings of Speech Prosody 9, Poznań, Poland.*

J. Michalsky, H. Schoormann, and O. Niebuhr. 2018b. Conversational quality is affected by and reflected in prosodic entrainment. In *Proceedings of Speech Prosody 9, Poznań, Poland.*

J. Ohala. 1983. Cross-language use of pitch. An ethological view. In *Phonetica*, 40: 1–18.

J. Ohala. 1984. An ethological perspective on common cross-language utilization of f0 in voice. In *Phonetica*, 41: 1–16.

D. P. Szameitat, K. Alter, A. J. Szameitat, D. Wildgruber, A. Sterr, and C. J. Darwin. 2009. Acoustic profiles of distinct emotional expressions in laughter. In *The Journal of the Acoustical Society of America* 126(1): 354-366.

J. Trouvain, and K. Truong. 2012. Convergence of laughter in conversational speech: effects of quantity, temporal alignment and imitation. In *Abstract ISICS 2012: International Symposium on Imitation and Convergence in Speech, Aix-en-Provence.*

K. Truong, and J. Trouvain. 2012a. Laughter annotations in conversational speech corpora – possibilities and limitations for phonetic analysis. In *Proc. 4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals, Istanbul*, pages 20-24.

K. Truong, and J. Trouvain. 2012b. On the acoustics of overlapping laughter in conversational speech. In *Proc. Interspeech, Portland.*

Cooperating with a smiling avatar: when face and voice matter

Ilaria Torre, Emma Carrigan, Killian McCabe, Rachel McDonnell,
Naomi Harte
Trinity College, Dublin

Abstract

Being able to express and interpret emotional expressions is paramount to a successful interaction. But what if the interlocutor expressing an emotion is a machine? The facilitation of human-machine communication and cooperation is of growing importance as smartphones, autonomous cars, or social robots increasingly pervade human social spaces. Previous research has shown that emotionally expressive avatars generally elicit higher cooperation and trust than 'neutral' ones (Elkins and Derrick, 2013; Krumhuber et al., 2007). Since emotional expressions are multi-modal, the question of which of the available channels should be most carefully considered in design arises. Would a mismatch in the emotion expressed in the face and voice influence people's cooperation with an avatar?

To answer this question, we developed a simulated survival game, originally devised by Hall and Watson (1970), where people had to cooperate with a computer-generated avatar in order to survive a crash landing on the moon. Participants' task was to rank a set of items in order of importance for survival, after the avatar gave some suggestions on how to rank them. How much participants accepted the avatar's suggestions was our implicit measure of cooperation (see e.g. De Houwer, 2006). The avatar's face and voice were designed to either smile or not, in 2 matched and 2 mismatched conditions: smiling voice and face, neutral voice and face, smiling voice only (neutral face), smiling face only (neutral voice). Thus, we could examine whether participants would cooperate with an avatar in one of these conditions more. The experiment was set up in a museum over the course of several weeks, where visitors were invited to interact with it.

Preliminary results from several hundreds visitors show that people tend to trust the avatar in the mismatched condition with the smiling face and neutral voice more. This suggests that participants might have found the avatar's smiling voice in particular to be off-putting. This has implications for Human-Machine Interaction and machine design.

References

Jan De Houwer. 2006. *What are implicit measures and why are we using them*. Thousand Oaks, CA.

Aaron C. Elkins and Douglas C. Derrick. 2013. The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents. *Group Decision and Negotiation*, 22(5):897–913.

Jay Hall and Wilfred Harvey Watson. 1970. The effects of a normative intervention on group decision-making performance. *Human relations*, 23(4):299–317.

Eva Krumhuber, Antony S. R. Manstead, Darren Cosker, Dave Marshall, Paul L. Rosin, and Arvid Kappas. 2007. Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, 7(4):730–735.

Audience Laughter Distribution in Live Stand-up Comedy

Vanessa Pope, Rebecca Stewart and Elaine Chew
Queen Mary University of London
School of Electronic Engineering and Computer Science
{v.c.pope, rebecca.stewart, elaine.chew}@qmul.ac.uk

ABSTRACT

Professional comedians are experts in the manipulation of group laughter, but how comedians manage group laughter live has yet to be explored. Stand-up comedy's repeated content across performances provides an opportunity to identify how performers design their content and delivery to control audience response, with possible applications to other interactional contexts. Using a novel stratified laughter representation, we describe the distribution of laughter types during 43 minutes of performance for an audience of approximately 150 people, then compare group laughter responses to the same comedy segment in five different performances. Frequent short bouts of laughter are present throughout the long-form performance.

We found more big audience laughter at the beginning of a performance and more small group laughter towards the end, suggesting that a comedian may be eliciting particular laughter types. When engaging the audience directly and deviating from the content of other performances, the comedian used self-laughter more frequently. Our findings suggest that a comedian's control of audience response is most visible in the relative timing of laughter. The same section of comedy material retained patterns in the gaps between laughter bouts across performances, showing that comedic timing may be as much about preventing laughter as eliciting it.

Keywords

Laughter; comedy; performance; timing; joke

1. INTRODUCTION

Stand-up comedians expertly deliver speech that makes their audiences laugh. Through an iterative process of rehearsal and performance, they have honed the craft of managing the responses of large groups of people. Performers repeat material show to show, offering a counterpoint to research into laughter in spontaneous conversation or in response to recorded media. Unlike research using video or audio of comedy to induce laughter in participants, comparative studies of live stand-up comedy can provide insight into how comedians manage laughter—or lack thereof—in a live interactive context.

In 1940, two experimenters used stopwatches and pencils to count laughter instances of one second or more across 13 performances of the same show [9]. While the number of laughs and their duration varied considerably, they found that the number of laughs correlated with audience size. Whether the same sections of performance elicited laughs was not noted, though the author states that "[one] exceptionally long laugh in the show continued for 18.4 seconds on its best night and 9.0 on its worst" [9](p. 183).

Group laughter has been studied using the ICSI dataset, comprised of recordings of meetings of six people on average [6, 12]. In the recorded project meetings, laughter accounts for 9% of vocalisation time, overlapping with other people's speech without the turn-taking associated with conversation [7]. In conversation, laughs most commonly come from the person who has just spoken, rather than as a response to humorous content [15], and most often not in response to formal attempts at humour [11]. Vettin and Todt found a median of 5.8 laugh bouts in 10 minute conversations, though the frequency and duration of laughter varied widely between participants [15]. In the ICSI corpus, laughter occurs on average once a minute [7]. For people watching comedy clips alone or with one other person, the average duration of a bout of laughter was under a second [1].

Studies of audience laughter have used its presence or absence in recordings (whether canned laughter or natural laughter) as a variable to examine how others' laughter affects participant perceptions of the speaker or the content (eg. [5, 2]). The contagious nature of laughter itself is often discussed [10, 3, 13], but whether laughter is timed consistently in response to the same material is less examined. The one-to-one relationship between humorous content and laughter has been questioned in research that has found that content can be laughed at before, during, and some time after the content itself [14]. Laughter has different forms and timings in conversational interactions [8, 4] and is likely to show the same variety in performative interactions. Predicting and managing these complex responses to humorous content is part of the performer's design process. Does an audience's response to comedy material support the idea that laughter is triggered by punchlines?

This research proposes a novel laughter representation and method, based on the estimated number of laughter participants, to investigate how audience laughter and performance interact, stratifying audience response and focusing on its timing and distribution.

2. METHODOLOGY

2.1 Data

2.1.1 *Performance Description*

The laughter data was recorded during the first act of a

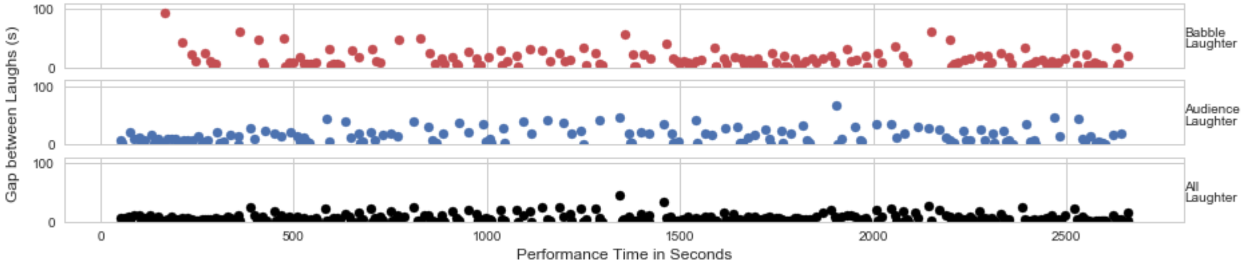


Figure 1: Gaps between laughter types in performance time in Show 3 (43 minutes)

solo professional stand-up comedy performer who had been on tour for over nine months. The stand-up comedian, SP, consented for their show to be recorded for research purposes. SP is British and their comedy style is conversational rather than consisting of a series of short jokes. Five performances were recorded between April and May 2017. All performances took place close to London with audiences of over 150 people.

2.1.2 Recording

Two types of recording were made using a Zoom H4N recorder during SP's performances to isolate the performer's voice from the audience noise. The Zoom recorder has two on-board microphones, in an XY setting at 90 or 120 degrees of one another, as well as having inputs to record from external microphones. To capture the cleanest audio possible of SP speaking, the performer's microphone feed was recorded directly from the sound desk. The Zoom's on-board microphones were placed on their widest stereo setting to record the ambient noise in the auditorium. The recordings analysed here are from the Zoom's on-board microphones.

Because the Zoom recorder was connected to the sound desk to access the microphone feed, its placement was determined by the venue. In four of the five venues, the sound desk was placed at the back of the auditorium, and in one performance the sound desk was backstage. During Show 3 the sound desk was at the back of the auditorium in an open booth with no glass or obstruction between the audience and the back row.

2.2 Annotation Labels

The initial goal of the annotation process was to identify how often laughter occurred during SP's performance to account for gaps in the timing of spoken material. However it quickly became apparent that a large proportion of audience laughter overlapped with the comedian's onstage speech. Focus shifted to the number of people involved in any given laughter bout to prepare for future work on how group laughter interacts with the performer's speech timing.

ELAN [16] was used to annotate the recording of audience laughter. The performer's voice can be heard in the background. The annotation labels used were:

Solo Laughter: Only one person laughing.

Babble Laughter: Between two and five people distinctly laughing.

Audience Laughter: More than five people laughing, or several people laughing in a manner that makes it hard to distinguish their number.

Self-laughter: The performer themselves laughing.

Applause or Cheers: Parts of performance where the audience clapped, whooped or responded with "oooo".

Audience, Babble and Solo Laughter are mutually exclusive categories.

3. RESULTS

To examine laughter distribution across performances, laughter types were first annotated for the entirety of one recording to select a section to examine in more detail across different performances. The first act of Show 3 was selected as it is the median performance; if the show is continually evolving at a similar rate, Show 3 is theoretically as different from the first recording as it is to the last.

3.1 Laughter Durations in Show 3

Table 1: Laughter Type Descriptions (Show 3)

	Audience	Babble	Solo	Self	All
Instances	158	166	86	10	420
Max. (s)	8.78	12.78	5.07	1.35	12.78
Min. (s)	0.73	0.50	0.29	0.32	0.29
Mean (s)	2.32	1.65	1.05	0.66	1.75
Total (s)	366.63	274.61	90.91	6.60	738.77

In Show 3, laughter of some kind was present for 27.5% of performance time. Audience Laughter accounts for slightly over half of total laughter time (Table 1). A similar amount of audience laughs and babble laughs were identified (158 and 166 respectively), but Babble Laughter was shorter on average, at 1.65 seconds compared to 2.32 seconds. Only 10 instances of self-laughter were identified. Laughter bouts lasted 1.80 seconds on average, reaching a maximum of 16.02 seconds.

3.2 Laughter Distribution in Show 3

Fig. 1 shows the gaps between laughter of each type. The longest gap with no laughter is 44.50 seconds and the average gap between laughs is 5.09 seconds. The distribution in performance time of each laughter instance shows that Audience and Babble Laughter are both frequent, but appear at different densities at different points in the performance.

Audience Laughter is more frequent at the beginning of the show (on the left), while there is a greater density of Babble Laughter towards the end (on the right). Comedians often begin a show with jokes about the venue or the audience itself, so this distribution may be showing the comedian using jokes that appeal to the audience as a whole

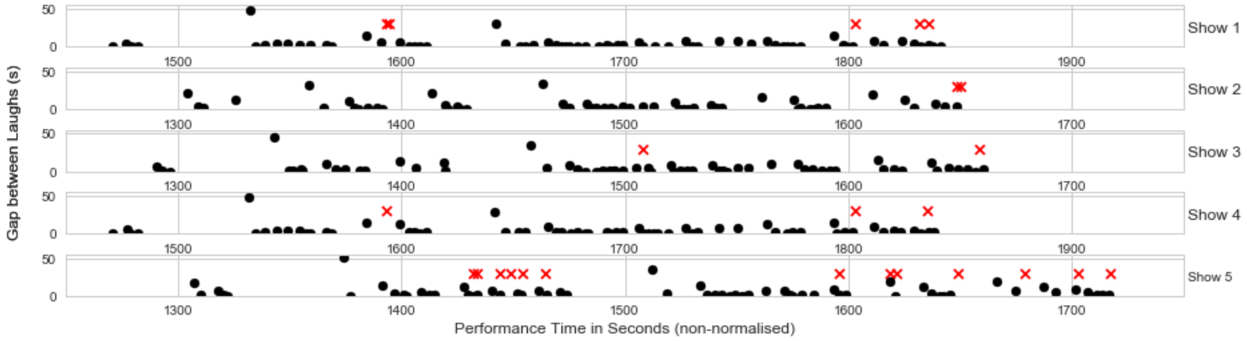


Figure 2: Gaps between group laughter bouts for the same segment of material in five different performances (6 minutes, approx.). Self-laughter start-times are marked with 'x'.

to win the audience over, gaining the goodwill necessary to build towards more complex jokes later in performance.

Considered alone, the average gap between laughs for Audience Laughter is 14.03 seconds and 14.13 seconds for Babble Laughter in Show 3. As expected, these gaps are more than double the average gap when all laughs are considered together. The gap between laughs drops when Audience and Babble Laughter are considered together, suggesting that Babble Laughter often occurs between instances of larger-scale audience laughter or vice versa.

Plotting the gaps between instances of Audience Laughter and Babble Laughter separately provides a more nuanced picture of how types of laughter are ebbing and flowing throughout the performance. A gap in laughter is measured as the duration between the end of one laughter bout and the beginning of the next, plotted at the point at which the second laughter bout began. Blank space followed by a low gap duration indicates that the blank space contained laughter, while blank space followed by a high gap duration indicates audience silence.

After the first 500 seconds (or roughly 8 minutes of performance time) frequent Audience Laughter, visible as a cluster of small gaps between laughs, gives way to a more even scattering of audience laughter. Babble Laughter becomes more frequent, particularly after the halfway mark. Examining the gaps between all audience laughter types together highlights parts of the performance where laughter was less frequent, visible as spikes in gaps between laughs.

3.3 Cross-performance Laughter Distribution

A short segment of material that contained the longest gap between group laughter in Show 3 (around 1250 seconds, see Fig. 1) was annotated in four other performances to see if the same material received a similar pattern of Audience Laughter. The roughly six-minute segment contained material around three themes: underwear, Brexit and Uber.

Across the performances the longest gap varied between 33.21 seconds (Show 2) and 50.39 seconds (Show 5), but it is still visible as significant, illustrating the importance of considering patterns in timing rather than absolute values. The average gap between laughs varied between 3.69 (Show 4) and 5.29 seconds (Show 4), though the standard deviation was more similar (between 6.92 in Show 1 and 8.78 in Show 5). Despite differences in statistical measures, long gaps between laughter instances appear at similar positions in performance time (Fig. 2). In some performances silences and frequent laughter alternate more clearly in binary patterns, such as in Show 1, whereas in other performances laughter gaps are more varied, such as in Show 5 where laughter gaps crest and drop. Rather than a punchline triggering a laughter type consistently, reliable audience response to content appears in gaps between sections of laughter, suggesting that performer control may be as much in holding back as in inducing audience response.

However, there is more variability in the duration of sections that contain many short burst of laughter. The long gaps between laughter instances contain very similar performance material, whereas there is more variability in the material eliciting short, frequent laughter. In Show 5, SP deviated from the material to engage the audience specifically about someone re-entering the room and to comment on their show. This new material is between the 1423 and 1455, visible in Fig. 2 as a section of short frequent laughter not present in other performances. Show 5's segment had 14 instances of self-laughter within this segment, more than the entirety of self-laughs identified in Show 3.

Here the comedian may be trying to spark audience laughter with their own laughter and additional material, trying to create a connection before moving on with the performance.

4. LIMITATIONS

Relying on only one microphone placement restricts the sources from which sound can be captured. The recorder was at the back of the room and audience members faced the front. The audio captured during the performance favours those sitting at the back of the room and may have missed quiet or solo laughter further away. Because of the potential to miss laughter, even faint or hard to make out solo laughter was counted to help redress the conservative laughter estimates this methodology produces. Microphone placement contributed to the decision not to use loudness of the laughter as one of the features to categorise laughter types.

The short duration of laughter bouts is partly due to the separation of laughter types. Babble Laughter often occurred close to Audience Laughter, while Solo Laughter occurred immediately before and after other laughter types as well as in isolation. Had they been annotated together, laughter bouts might be considered longer and less frequent. The distinct acoustics and differing distributions of laughter types suggest that combining all group laughter may miss subtle distinctions in group laughter dynamics.

5. DISCUSSION

Despite the conversational style of SP's comedy, it seems that they have some control of when the audience responds in this one-to-many interaction. The frequency of laughter overall suggests that sections with infrequent laughter may be significant. Parts of a mature show in which the audience is consistently quiet are likely to be purposefully placed, sculpting audience attention by offering a counterpoint of comparatively serious silence. Orchestrated group silence may be as powerful an indicator of performer planning as group laughter.

Future work on the dynamics of group laughter in audiences, and the performer's control or reaction to it, will require a fine-grained approach to timing given the frequency and short durations of laughter bouts. Our annotation schema does not capture the differences in laughter texture that influenced interactions: some audience laughter was muted, underscoring the performer's speech, while other audience laughter was loud and synchronous and occurred in pauses in speech. It would be interesting to explore whether the performer is also controlling the type of laugh, encouraging supportive underscoring laughter or disruptive responses from the audience as part of their performance's design. In Show 5 self-laughter was used alongside extra material that seemed aimed to stimulate audience engagement between signature laughter gaps, suggesting that the performer may have been trying to get the audience in a particular state before beginning the next section of prepared material.

This case-study of audience laughter types will be extended to examine a section of material from a stand-up comedy routine performed to 10 different audiences. Comparing patterns of audience laughter, and whether they coincide with the same pieces of prepared material, would illuminate the extent to which the performer controls audience laughter through the design of their material or spontaneously in their delivery. Particular attention will be given to patterns in audience response, joke consistency and the introduction of new material.

6. CONCLUSION

Stand-up comedy audiences laughed during 27.5% of a 43 minute performance with frequent, short laughter bouts. The frequency of laughter makes gaps in laughter of interest. Comparing laughter distribution in a segment of material present in five comedy performances showed that gaps between audience laughter were maintained across performances, even as the frequency and duration of laughter varied. The performer's self-laughter appeared alongside new material aimed at audience engagement, suggesting that the performer is actively manipulating how their audience laughs. Despite the fact that laughter was found before, during and after punchlines, there were consistent sections of performance with no laughter at all.

7. ACKNOWLEDGMENTS

The authors would like to thank comedian SP for their participation. This research was supported in part by the Media and Arts Technology Centre for Doctoral Training, funded by EPSRC and AHRC (ref EP/L01632X/1).

8. REFERENCES

[1] J. A. Bachorowski, M. J. Smoski, and M. J. Owren. The acoustic features of human laughter. *The Journal of the Acoustical Society of America*, 110(3):1581–1597, 2001.

[2] A. J. Chapman. Funniness of jokes, canned laughter and recall performance. *Sociometry*, 36(4):569–578, 1973.

[3] A. J. Chapman and W. A. Chapman. Responsiveness to humor: Its dependency upon a companion's humorous smiling and laughter. 88(2):245–252, 1974.

[4] J. Ginzburg, E. Breithholtz, R. Cooper, J. Hough, and Y. Tian. Understanding laughter. In *Proceedings of the 20th Amsterdam Colloquium*, 2015.

[5] C. R. Gruner. Audience's response to jokes in speeches with and without recorded laughs. 73(1):347–350, 1993.

[6] A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, and A. Stolcke. The ICSI meeting corpus. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 1, pages I–I. IEEE, 2003.

[7] K. Laskowski and S. Burger. Analysis of the occurrence of laughter in meetings. In *Eighth Annual Conference of the International Speech Communication Association*, 2007.

[8] C. Mazzocconi, Y. Tian, and J. Ginzburg. Multi-layered analysis of laughter. *SEMDIAL 2016 JerSem*, 2016.

[9] J. Morrison. A note concerning investigations on the constancy of audience laughter. 3(2):179–185, 1940.

[10] R. R. Provine. Contagious laughter: Laughter is a sufficient stimulus for laughs and smiles. 30(1):1–4, 1992.

[11] R. R. Provine. Laughter punctuates speech: Linguistic, social and gender contexts of laughter. 95(4):291–298, 1993.

[12] E. Shriberg, R. Dhillon, S. Bhagat, J. Ang, and H. Carvey. The ICSI meeting recorder dialog act (MRDA) corpus, 2004.

[13] M. M. Smyth and R. G. Fuller. Effects of group laughter on responses to humorous material. 30(1):132–134, 1972.

[14] Y. Tian, C. Mazzocconi, and J. Ginzburg. When do we laugh? In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 360–369, 2016.

[15] J. Vettin and D. Todt. Laughter in conversation: Features of occurrence and acoustic structure. 28(2):93–115, 2004.

[16] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes. ELAN: a professional framework for multimodality research. In *5th International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1556–1559, 2006.

Sighs in everyday and political communication

Isabella Poggi, Alessandro Ansani, Christian Cecconi

Department of Philosophy, Communication and Performing Arts
Università Roma Tre

Abstract

The work defines the sigh as a type of breath expressing or communicating specific mental or emotional states. To investigate the meanings of the sigh, after overviewing preliminary analyses of written and oral corpora, the paper focuses on a peculiar use of it as a "discrediting body comment" often exploited in political debates to imply the opponent's stupidity or obsessive repetition, by displaying frustration or boredom. In a perception study on some uses of this sigh, participants' interpretations are not significantly shared, but agreement emerges when considering their perception of sighs in terms of valence and arousal.

1 Introduction

Among the types of vocalization produced in vocal communication, some exploit the potentialities of breath for expressive or communicative aims. A snort may be issued when climbing steep stairs to convey effort, voiced expiration might communicate "I'm fed up", a sudden voiced inspiration is often part of a startle reflex.

This work investigates the meanings of the sigh, a type of breath that can be used by a person both when alone – thus conveying expressive contents – and addressed to other people, with or without a communicative goal. Sect. 2 presents related work on sighs, Sect. 3 proposes a definition of sigh and overviews preliminary works on its meanings, Sect. 4 presents a perception study on a particular use of sighs in political debates.

2 Related work on sighs

Sighs have been studied in terms of their physiological mechanism and of their semantic and interactional functions.

Physiologically, a sigh is triggered by a normal eupneic breath and followed by a respiratory pause, a "post-sigh apnea" (Ramirez, 2014); it is a deep breath, a second breath starting when another is not yet over (Boncinelli, 2016; Li et al., 2016), that due to actual need for more air or to emotional events triggers a second inspiration.

In a Conversation *Analy*sis framework, analyzing pre-utterance, post-utterance, stand-alone and transitional sighs, Hoey (2014) attributes them various functions and meanings depending on their position, context, physical production. Being manipulable, sighs can be conscious and used for social ends, and when and how they are delivered may influence people's perception of their meaning in social interaction.

On the psychological side, Teigen (2008) investigated sighing in three works: a survey showed that people associate sighing mainly to negative, low-intensity and low-arousal emotional states; then it was shown that sighs produced by others are widely attributed to sadness, while self-produced sighs are mostly interpreted as "giving up" or "surrendering". In a study asking participants to solve impossible puzzles, the sighs elicited by their ineffective attempts were seen as mostly unintentional expressions of some course of action, a wish, a plan to be set aside, a pause before a plan is replaced by a newfound initiative. Teigen (2008) finally proposes the following list of sigh types:

(1) Sadness (including sorrow, depression, disappointment, and loneliness)
(2) Giving up (resigned, helpless, despondent)
(3) Weariness (tired, exhausted)
(4) Boredom (unmotivated, restless)
(5) Frustration (stress, irritated, displeased)
(6)Other negative emotions (jealous, afraid,

nervous, envious, hungry)
(7) Happiness (joy, in high spirits, in love)
(8) Satisfaction (relieved, well-being, content)
(9) Relaxed (silent, tranquil)
(10) Empathy (sympathetic, compassionate)
(11) Other (surprised, excited, "strong feelings")

3 The meanings of sighs: a first overview

This work focuses on a particular use of sigh in multimodal political debates. But before going into this, let us define the sigh in terms of a socio-cognitive model of communication (Poggi, 2007). A sigh is a vocal signal that may have either an expressive or a communicative function: it is expressive when its breathing pattern simply displays some internal physical or mental states like weariness or sadness, without its Sender having a conscious goal of making others know about such internal state. A sigh is communicative, instead, when its Sender has the conscious goal of having another know about his internal state. Whether expressive or communicative, it is a holophrastic signal (Poggi, 2009), i.e., it conveys a whole communicative act, including a performative and a content, where the performative is one of information and the content is an internal physical or psychological state. Thus the semantic structure of a sigh is always: "I inform you I am feeling X" where X may assume the meaning of various possible internal states, e.g. those overviewed by Teigen.

Before focusing on a particular type of sighs, we conducted two preliminary studies. First we explored the polysemy of sighs in a qualitative study on written literary texts: searching for the root *sospir-* (= sigh in Italian) in a corpus of 97 occurrences from 37 novels, for each occurrence we provided a verbal paraphrase of the meaning conveyed in that context, finding out, like did Teigen, that sighs convey: physical states (tired, exhausted, weary); negative emotions and mental states (sadness, being fed-up, grief, displeasure, regret, resignation, giving up); positive mental and emotional states (desire, patience, relief.

In a second study, we collected a corpus of 100 videos, taken from movies, tv fiction, cartoons, talk shows, political debates, where characters or debaters sigh during interaction. In this corpus too we generally found the same meanings as Teigen (2008); yet two peculiarities popped up. First, we also found a positive meaning of "self-encouragement", in case of preparation for an effort, either physical or mental. Second, we found that in most political debates the sigh is a body comment aimed at discrediting, to the point of delegitimizing, the opponent by implying s/he is boring or stupid. See this "discrediting sigh".

(1) https://www.youtube.com/watch?v=X2XPD4Y6gs4
Laura Boldrini, the leftist Chair of the Italian Chamber, while talking of the boat people arriving on the Italian coasts, argues against her present opponent Matteo Salvini, the rightist leader of the North League, who claims the necessity to push them back, that this situation is the fault of the previous policy of the right government. While she is talking, Salvini performs an *audible inspiration* while *rolling his eyes up*, then he *points his eyes* again *to the camera* with his *eyelids half open*, making an *audible expiration*.
This sigh looks as a signal of impatience and intolerance addressed to Boldrini's complaint.

This type of sigh, working as a flaunted expression of annoyance and intolerance, is often exploited in political debates as a "discrediting body comment" (Poggi et al., 2012), a way to express one's negative evaluation of the opponent's discourse by facial expressions or other body signals that provide a "silent" feedback to the audience during the present speaker's turn. When politician A is talking, the opponent B must leave him/her the turn, but taking advantage of being video-recorded by the camera, s/he launches seemingly "silent" messages to the audience, thus implicitly or explicitly displaying disproval through expressions of boredom or annoyance, for instance by *rolling eyes*, *looking up in the sky*, or just *sighing*. Generally, the sigh has a literal meaning of frustration or boredom. Eyes upward, rolling eyes, opening arms may or not cooccur with the sigh, but when they do they enhance its aggressive import, by their very meaning, a pretended prayer to God. In this case, the eye signals intensify the expression of negative emotion, as if saying: "I am frustrated / bored", "God, I pray you (help me bear this)". The emotion display, whether intensified or not, in its turn implies negative evaluations about the opponent: expressing frustration may imply the other (or his discourse) is so stupid as not to be amendable; expressing boredom implies he/it is repetitive or pointless.

To investigate the meanings of sighs in general, and of their peculiar "discrediting sigh" among others, we conducted a perception study.

4 Sighs: a perception study

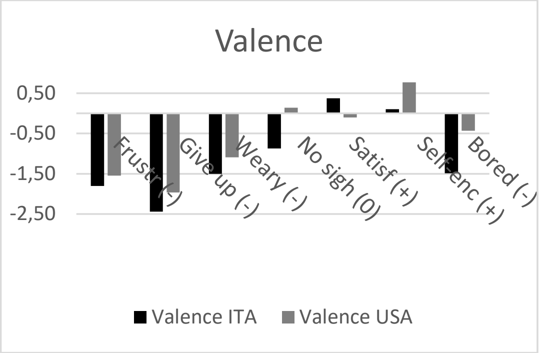
The goal of our study was to check if people viewing and listening to different sighs can attribute them different meanings, and if these meanings are shared among judges in the particular context of political debates.

In order to a preliminary check of Teigen's (2008) taxonomy of meanings in the context of political debates, and to select the subset of meanings to submit to participants in our study, 55 different sighs, all taken from Italian political tv-shows, were analysed by two independent judges. Within these, to better adapt the list to interaction in the political context, that is generally not so placid or relaxed, we selected only 7 items: 5 out of them correspond to Teigen's categories, while we excluded those with a positive valence (happiness and empathy) and, for balancing reasons, his category of *other negative emotions* (jealous, afraid, nervous, envious, hungry). Instead, we included the new positive category of self-encouragement found in the preliminary corpus analysis, to assess, through our perception study, if this is actually a possible meaning of the sigh. Finally, as a control item we included a case in which no sigh actually was produced.

Thus the selected items, beside the no-sigh, included frustration, boredom, weariness, giving up, satisfaction, self-encouragement: states with different combinations of valence and arousal. Some items were deliberately ambiguous among 2 meanings.

4.1.Participants and method

64 participants, 34 from USA and 30 from Italy were recruited through an online campaign and submitted with a survey in which they had to watch 8 different videos taken from Italian political debates, 6 of which contained a sigh. The small number of items was aimed at preventing overload in participants. The videos of political debates were selected based on the technical feature that both opponents were visible in simultaneous frames, so one could see both the politician presently speaking and the one displaying facial comments during the other's turn. Participants from the USA were asked to rate their level of understanding of Italian. After watching each video, both Americans and Italians were presented with the whole list of Teigen's meanings, with the addition of "self-encouragement", and they were asked to tell, on a 7-points Likert scale, how much each of those meanings corresponded to the sigh in the video.



4.2. Results and discussion

As results from Table 1, the sighs of frustration, self-encouragement and boredom were quite frequently recognized as such by participants, while those for giving-up, weariness and satisfaction elicited sparser ratings. The regression toward the mean for the control item reveals that participants can tell the difference between what is a sigh and what is not.

		Stimuli					
	Frustr	Give-up	Weary	Satisf	Self-enc	Bored	No sigh
Sadness	2,36	2,58	2,28	2,16	1,89	2,30	2,39
Giving up	3,91	4,23	3,69	2,66	2,50	2,97	2,91
Weariness	3,61	4,25	3,80	2,70	2,58	3,38	2,81
Boredom	3,88	4,20	**4,05**	2,27	2,56	**4,06**	2,67
Frustration	**4,64**	**5,44**	4,02	**3,66**	2,78	3,20	2,77
Satisfaction	1,70	1,67	2,27	2,77	2,77	2,28	2,23
Relaxed	1,80	1,73	2,30	2,30	3,11	2,55	2,20
Self-enc.	2,41	2,25	2,44	3,45	**3,38**	2,23	2,72
Other	2,14	2,16	2,13	2,72	2,44	1,95	2,33

Table 1. Ratings of sighs

Moreover, also when the specific emotion is not recognized, participants generally correctly rate the sigh either in terms of the dimension of valence (e.g., frustration perceived as giving up), or in terms of arousal (e.g., boredom perceived as relaxation). In general sighs conveying negative valence are perceived in any case as negative, but slightly more so by Italians than Americans (positive valence Italians mean = 0,24; positive valence Americans mean = 0,34; negative valence Italians mean = -1,81; negative valence Americans mean = -1,26). In Fig.1, on the x axis we labelled each video as +, - or 0 depending on our hypothesis on the relative valence conveyed by its sigh; on the y axis we listed the mean of a composite score [(satisfaction + relaxation + self-encouragement + other/surprised/excited) – (sadness + giving up + weariness + boredom)] showing the level of valence indicated by participants; as can be seen, they answered coherently with our hypothesis.

Figure 1: Valence among all participants

We think that a more frequent attribution of negative valence to sighs in the videos on the part of Italian participants might be due not only to mere language competence, but to cultural knowledge in a broad sense: in most videos the Senders of the sighs were some politicians (e.g., Matteo Salvini) or journalists (Marco Travaglio) that are known to be particularly sarcastic. To this we might add also possible personal and political sympathy of Italian participants toward the speaker or the "sigher" in the debate, that might have influenced their interpretation of the sigh meanings, by viewing them as more or less aggressive than Americans did.
As regards the self-encouragement sigh, Italian participants generally tended to recognize it; the same cannot be said for Americans, who preferably rate it as conveying relaxation and satisfaction [Fig. 2].

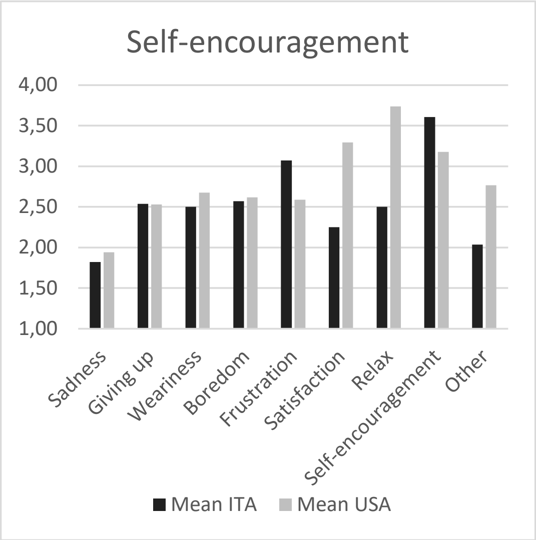


Figure 2: Self-encouragement

4 Conclusion and future work

The sigh is a highly polysemic signal, and its multiple meanings are not easy to distinguish; yet participants in our perception study differentiate positive from negative, and high from low arousal sighs.

Future research will try, first, to set a clear distinction, in terms of perceivable features and meanings, between sighing and other vocalizations like panting, puffing or snorting. Then, the correspondences will be investigated between the physiological production of sighs and consequent perceivable features and their respective meanings, checking for instance whether the audibility of inspiration and expiration correlates with different interpretations. Finally, it will be further investigated what meanings are added by the combination of a sigh with other body signals, such as rolling eyes, opening arms, raising head. The simulation of sighs in Virtual Agents will be both an end and a tool for such investigation.

Acknowledgments

The work is partially funded by the Italian National Project PRIN "Cultural Heritage Resources Orienting Multimodal Experiences (CHROME)" n. 2015WXBPYKCHROME.

References

Boncinelli E. 2016. Il sospiro, estenuato anelito di vita. *Corriere della Sera*, March 27, 2016.

Hoey E.M. Sighing in interaction. 2014. Somatic, Semiotic, and Social. *Research on Language in Social Interaction*, 47(2): 175-200. https://dx.doi.org/10.1080/08351813.2014.900229.

Li P. and Yackle K. 2017. Sighing. *Current Biology* 27, February 6, 83-102.

Poggi I. 2007. *Mind, hands, face and body. A goal and belief view of multimodal communication*. Berlin, Weidler.

Poggi I. 2009. The language of Interjections. In A.Esposito, A.Hussai, M.Marinaro & R.Martone (Eds.) *Multimodal Signals: Cognitive and Algorithmic Issues*. Berlin, Springer, pp.170-186.

Poggi I., D'Errico F., and Vincze L. 2013. Comments in words, face and body. *Journal of multimodal user interface*, 7 (1): 67-78. DOI: 10.1007/s12193-012-0102-z.

Ramirez JM. 2014, The integrative role of the sigh in psychology, physiology, pathology, and neurobiology. *Prog. Brain. Res.*, 209, 91-129.

Teigen K.H. 2008. Is a sigh "just a sigh"? Sighs as emotional signals and responses to a difficult task. *Scandinavian Journal of Psychology*, 49, 49–57.

Annotating Nonverbal Conversation Expressions in Interaction Datasets

Kevin El Haddad, Noe tits, Thierry Dutoit
numediart Institute, University of Mons /31 Boulevard Dolez, Mons, Belgium
{kevin.elhaddad, noe.tits, thierry.dutoit}@umons.ac.be

Abstract

In this paper, we present our work on building a database of Nonverbal Conversation Expressions (NCE). In this study, these NCE consist of smiles, laughs, head and eyebrow movements. We describe our annotation scheme and explain our choises. We finally give inter-rater agreement results on small part of the dataset

1 Introduction

Virtual agent systems like chatbots, virtual assistants, etc. have seen a lot of improvements in the last decades thanks mainly to the progress of artifical intelligence in general and machine learning/deep learning in particular. These systems are becoming more and more part of our daily lives and will become more enchored in it in the near future. It it therefore important that our interactions with them be as comfortable as possible. This is why it is important for them to better understand the human ways of interaction and also to be able to behave in a human-like way.

Nonverbal and paralinguistic expressions form a big part of human-human interactions. They are very frequent and have different important functionalities. It was reported that laughter, for instance, accounted for about 10% of the total verbalizing time(**?**). Other studies also report the importances of these nonverbal expressions in interactions(). But they are yet to be well implemented in human-agent interaction systems.

In this paper we present an ongoing work on building a nonverbal conversation expression dataset. Nonverbal conversation expressions or NCE (El Haddad, 2017) are expressions that come to complement the semantic of a sentence's linguistic content (e.g. emotional speech), or as standalone expressions that are understandable without needing words (e.g. nodding, smiling, affect bursts, etc...) .

The main purpose of the database is to be used to build human-agent interaction systems. Considering the efficiency of artificical intelligence in general and deep learning in particular, the dabase should be oriented, among other things, to deep learning applications.

2 Data Used

In order to answer deep learning systems needs, the ulitmate goal of this work is to obtain a large database of NCE. So the work presented here should be applied on different open-source and available databases of interactions. However, for now, we are using a dataset comprising audio and video recordings of dyadic conversations for which the topic was moral emotions (Heron et al., 2018). Moral emotions are emotions that are ethically relevant (Haidt, 2003) such as (gratitude, aw, empathy, shame, etc...). The setup of this dataset was made in a way to control the listener/speaker roles. Each of the participants was assigned randomly the role of the speaker or listener. The listener was told to ask the speaker predefined questions about moral emotions in the form *"When was the last time you felt ...?"*. The moral emotions in question were: shame, guilt, compassion and gratitude. Then the speaker/listener roles and questions were altered randomly until the questions for all emotions were asked. This way, the dataset provides data of speaker and listener expressions during a naturalistic interaction. The dataset contains 21 sessions (42 speakers) of 14 different nationalities. Each session containing 4 topics, one for each emotion asked. It is worth noting that due to the hardware setup (microphones and cameras) the data contain overlapping speech.

3 NCE Annotation

Intuition

As mentioned previously, the goal of this database it to help building human-agent interaction systems. Therefore, we consider that the data should be useful mainly for detection systems, decision making and generative systems.

So the annotations undertaken here will focus on localizing the start and end times of different NCE as accurately as possible. In this work, the functionality of the NCE are not considered. We consider only the event independently from the social function, intend/purpose of the expressions, situation or context. Two main reasons are behind this choice.

1. Annotating such contextual information would be a lot more challenging, tedious and time consuming than just delimitting the event. Indeed the values to be considered must be decided beforehand and more time will be required for each annotation. Also, such expressions might be dependent on the individual's culture, personality and even on the state of mind at the time of recording. Which are information for which the access is difficult and sometimes impossible especially if our ultimate goal is to obtain enough data for machine learning and deep learning systems.

2. Deep learning systems have already shown their ability to learn internal representations of the data and the task. So we hope that, with enough data such systems can be used to map specific NCE with specific context, situations and subject without requiring such annotation task.

With the NCE time intervals we will be able to train supervised machine-learning classifiers, build expression prediction systems for speakers/listeners and synthesis-by-concatenation systems like in (El Haddad et al., 2016b) and even audiovisual generative systems.

Annotation Scheme

Based on the literature related to several NCE, we consider, here, 4 different NCE in this work: smiles, laughter, head and eyebrow gestures. The criteria we used for this choice is are the fact that: i) they occur frequently in human-human interactions ii) they play a role in dialog strategies and phenomena like mirroring.

Expression	Values
Smiles	subtle, low, medium, high
Laughs	low, medium, high
Head movements	nod, shake, tilt
Eyebrow gestures	left/right/both raise/frown

Table 1: NCE annotation values

Indeed it has been shown in several previous separate work that these 4 expressions answer both of these criteria by happening frequently in dialogs and by being used for mirroring and other functionalities (Paggio and Navarretta, 2011b; Navarretta, 2016; Paggio and Navarretta, 2011a; Aubrey et al., 2013; McKeown et al., 2012; Dupont et al., 2016; Paggio and Navarretta, 2017; El Haddad et al., 2016a).

Each of the above-mentioned expressions will have descriptive values as shown in Table 1 and as detailed in what follows.

Smiles and Laughter: Both of these expressions have been the subject of many studies (El Haddad et al., 2016a). intensity or arousal is very important for both of these expressions. Indeed, in (McKeown and Curran, 2015) presents a study the relationship between laughter intensity and humor.

Concerning the smiles, the definition we are using is not focused on the lips movements alone. Several studies of the smile facial expressions can be found. Most of them agree that the Action Units (AU) corresponding to cheek raising (AU06) and lips spreading (AU12) respectively are important to consider (Ochs et al., 2017; Ekman and Friesen, 1982). But also lower eyelids raised (AU7), lips upside down (AU15) or pressing the lips together (AU24) have also been reported to be linked to smiling. But smiles can occur while speaking or while doing other facial expressions, for example, compressed smiles can be a combination of lips spreading (AU12) with turning the lips upside down (AU15) or pressing the lips together (AU24) (Harris and Alvarado, 2005; Ekman and Friesen, 1982). These facial expressions will therefore be used to determine the occurrence or not of a smile. Then, the smiles are segmented based on their intensity levels. The intensity is itself based on the intensity of the facial expressions used to deter-

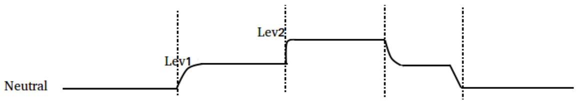


Figure 1: Example of segmentation of different level of smiles based on the intensity levels.

mined it was a smile. We define three different intensity levels (low, medium and high). One of the particularities of our annotation scheme is that we consider the smiles of very low level that seem to last ”all the time”. Chovile did not consider smiles in (Chovil, 1991), as smiles were so overwhelmingly frequently present in the data compared to other expressions. Similarly, many databases neglect these types of smiles. We decided to annotate them because they are part of the interaction and must have an effect since they can be perceived. So, we include a fourth level too: subtle (not related to the term used for micro-expressions). This is to annotated smiles of very low intensity which usually stay for a long period of time (and sometimes not) and to which it is sometimes hard to associate a specific AU or facial expression.

In order to have precise limits between two smiles, we rely on the transitions. Indeed, the work presented in (Schmidt et al., 2003) shows the importance of the speed of the transition from one expression to another. The choice of the intensity is somewhat subjective. The segments will start and end at the beginning of a level and the beginning of the next level respectively. An example is shown in Fig. 1

For laughs, the segments start when an audio, facial expression or body movement related to laughter is observed and stops when a breath intake is perceived whether audibly or visually (from the stomach, face, etc.). If no breath intake is perceived the end of the segment is considered to be when the movement stops.

Finally, we consider that these laughter and smiles cannot overlap: a laughter is not a smile and a smile with one of the movements mentioned above is a laugh.

Head and Eyebrow Movements: For head movements we consider nodding, shaking and tilting: pitch, yaw and roll movements respectively. The segments start and end with the movements. In the case of tilting, the annotations do not include the static head bent on the side after the movement has occurred. Only the movement is annotated. Considering the eyebrow movements, we annotate the raise and frown states of each or both eyebrows. Unlike the head movements, the annotations are not based on the movement only. The segments start when the movement starts and ends when the eyebrow is not perceived as raised or frowned anymore, taking the raised or frown state in between into account.

NCE	**Smiles**	**Laughs**	**HM**	**EM**	**All**
CKC	0.47	0.43	0.48	0.15	0.4

Table 2: Cohen’s Kappa Coefficients (CKC) to estimate inter-rater agreement

4 Inter-rater Agreement

Until now, 27 topics (part of a session) are annotated for the speaker and the corresponding listening in the dataset mentioned above, only 4 of which were annotated by 2 annotators. The total amount of time of 7 minutes and 11 seconds of data. Fig. 2 show examples of the obtained results for smiles, laughter, head and eyebrow movements for each of the annotators with respect to time. The integer values on the ordinate axis correspond to the intensity levels in case of the smiles and laughs (the lower the integer the lower the intensity (0 corresponding to neutral). In the case of head movements they correspond to nod (1), shake (2), tilt (3) and no movement (0). In the case of eyebrow movements 1 corresponds to raised (whether it is both eyebrows or only one), 2 to frown (none in this case) and 0 to no movement. The Cohen’s Kappa Coefficients were calculated to estimate the inter-rater agreement. The results mean values are given in Table 2.

Considering the complexity of the choice making and that part of the annotations were rather subjective an average Cohen’s Kappa of 0.4 is acceptable.

5 Future Work

After the dataset mentioned here is fully annotated we intend to use it to build NCE detection, prediction and generation systems. We also intend to carry on the annotations to other datasets as well.

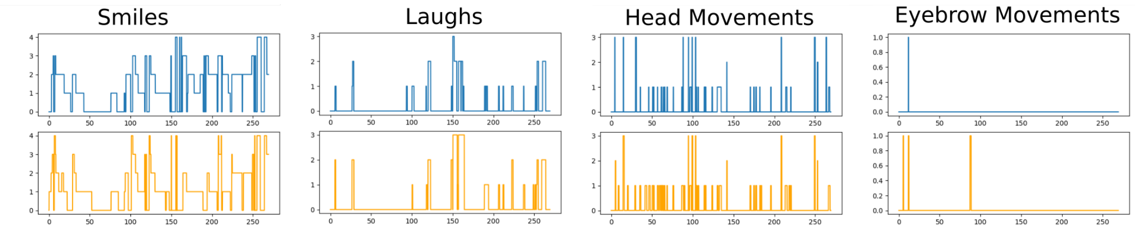


Figure 2: Annotations with respect to time for 2 annotators (blue and orange). The integers (1 to 4) correspond to the different annotation values corresponding to each expression mentioned in Table 1.

References

Andrew J Aubrey, David Marshall, Paul L Rosin, Jason Vandeventer, Douglas W Cunningham, and Christian Wallraven. 2013. Cardiff conversation database (ccdb): A database of natural dyadic conversations. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 277–282. IEEE.

Nicole Chovil. 1991. Discourse-oriented facial displays in conversation. *Research on Language & Social Interaction*, 25(1-4):163–194.

Stéphane Dupont, Hüseyin Çakmak, Will Curran, Thierry Dutoit, Jennifer Hofmann, Gary McKeown, Olivier Pietquin, Tracey Platt, Willibald Ruch, and Jérôme Urbain. 2016. Laughter research: a review of the ilhaire project. In *Toward Robotic Socially Believable Behaving Systems-Volume I*, pages 147–181. Springer.

Paul Ekman and Wallace V. Friesen. 1982. Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, 6(4):238–252.

Kevin El Haddad. 2017. Nonverbal conversation expressions processing for human-agent interactions. In *Affective Computing and Intelligent Interaction (ACII), 2017 Seventh International Conference on*, pages 601–605. IEEE.

Kevin El Haddad, Hüseyin Cakmak, Stéphane Dupont, and Thierry Dutoit. 2016a. Laughter and Smile Processing for Human-Computer Interactions. In *Just talking - casual talk among humans and machines*, Portoroz, Slovenia.

Kevin El Haddad, Hüseyin Çakmak, Emer Gilmartin, Stéphane Dupont, and Thierry Dutoit. 2016b. Towards a listening agent: A system generating audiovisual laughs and smiles to show interest. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI 2016, pages 248–255, New York, NY, USA. ACM.

Jonathan Haidt. 2003. The moral emotions. handbook of affective sciences.

Christine Harris and Nancy Alvarado. 2005. Facial expressions, smile types, and self-report during humour, tickle, and pain. *Cognition and Emotion*, 19(5):655–669.

Louise Heron, Jaebok Kim, Minha Lee, Kevin El Haddad, Stephane Dupont, Thierry Dutoit, and Khiet Truong. 2018. A dyadic conversation dataset on moral emotions. In *Proceedings of the Workshop on Large-scale Emotion Recognition and Analysis, Face and Gesture*, China.

Gary McKeown, Roddy Cowie, Will Curran, Willibald Ruch, and Ellen Douglas-Cowie. 2012. Ilhaire laughter database. In *Proceedings of 4th International Workshop on Corpora for Research on Emotion, Sentiment & Social Signals, LREC*, pages 32–35. Citeseer.

Gary McKeown and Will Curran. 2015. The relationship between laughter intensity and perceived humour. In *The 4th Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisations in Speech, Enschede, Netherlands*, pages 27–29.

Costanza Navarretta. 2016. Mirroring facial expressions and emotions in dyadic conversations. In *LREC*.

Magalie Ochs, Catherine Pelachaud, and Gary Mckeown. 2017. A user perception–based approach to create smiling embodied conversational agents. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(1):4.

Patrizia Paggio and Costanza Navarretta. 2011a. Feedback and gestural behaviour in a conversational corpus of danish.

Patrizia Paggio and Costanza Navarretta. 2011b. Head movements, facial expressions and feedback in danish first encounters interactions: A culture-specific analysis. In *Universal Access in Human-Computer Interaction. Users Diversity*, pages 583–590, Berlin, Heidelberg. Springer Berlin Heidelberg.

Patrizia Paggio and Costanza Navarretta. 2017. The danish nomco corpus: multimodal interaction in first acquaintance conversations. *Language Resources and Evaluation*, 51(2):463–494.

Karen L Schmidt, Jeffrey F Cohn, and Yingli Tian. 2003. Signal characteristics of spontaneous facial expressions: Automatic movement in solitary and social smiles. *Biological Psychology*, 65(1):49–66.

The MULAI Corpus: Multimodal Recordings of Spontaneous Laughter in Dyadic Interaction

Michel-Pierre Jansen[1], Dirk K.J. Heylen[1], Khiet Truong[1], Gwenn Englebienne[1] and Deniece S. Nazareth[1,2]
[1] Human Media Interaction department - University of Twente
[2] Department of Behavioral, Management and Social Sciences - University of Twente
For correspondence please email the first author: m.jansen-1@utwente.nl

Abstract

We present the MULAI Database that aims to provide researchers with more data to study the expressive patterns that humans demonstrate while laughing during human-human interactions. We collected a multimodal database that contains 357 minutes of recorded video-, audio-, and physiological data on dyadic interactions that often was paired with laughter. In addition personality questionnaire data was retrieved from all the participants. The database is unique in that it is recorded in several modalities not often explored and includes both spontaneous- and task induced laughter.

1 Introduction

Although many theories exist about the exact origin and function of laughter (Gervais and Wilson, 2005), most theorists believe laughter also plays a major social role. Laughter is one of the most common communicative signals. However, despite this fact knowledge about the multimodal expressive patterns is still rather limited (Niewiadomski et al., 2013). Therefore there is a need for specific data to study the multimodal expressive patterns of laughter. Even though there are now several databases available that address this need for information, there are still important gaps that need to be filled.

In this paper we introduce a new multimodal database focused on capturing expressions of laughter through the use of recording equipment that can obtain high quality audio-, high definition video- and physiological data. We will first briefly review a non-exhaustive list of available databases. Next we will describe the current database in full detail. In the end we will discuss future research that we are planning to do with this database.

2 Related databases

Table 1 provides an overview of some of the relevant databases of the last decade that contribute to laughter research, for more details we suggest the works of Petridis et al. (2013); Niewiadomski et al. (2013).

In general databases have different contexts in which laughter is elicited, the contexts in which no interaction (single participant), dyadic interaction or multi-party interaction occurs seems to be most prevalent.

Another way to sort the databases is through used elicitation methods for eliciting laughter. This includes spontaneous laughter during free conversations, induced laughter through the participation in ”funny” tasks, watching video and other stimuli that induce laughter or posed laughter.

From observing table 1, it can be deduced that not many modalities outside of the audio- and visual modalities are explored. Only the MMLI (Niewiadomski et al., 2013) and in some form the MANHOB (Petridis et al., 2013) database provide other modalities to analyze when researching laughter. This database aims at providing multiple modalities outside the normal range of modalities. In addition it provides a mixture interaction based laughter elicited through tasks or occurring spontaneously in non-task related conversation.

3 The MULAI Database

The MULAI database contains 13 sessions of dyadic interaction and laughter with multiple video-, audio- and physiological data streams recorded. In total 357 minutes of both video- and separate audio clips were obtained, containing both spontaneous and task induced laughter.

Database	Context of Laughter	Elicitation method	Modalities	Reference
MULAI	Dyadic int.	Spontaneous, task	BM, ECG, GSR	-
AMI	Multi-party int.	Spontaneous		Mccowan et al. (2005)
AVLC	Single part.	Video		Urbain et al. (2010)
AVIC	Dyadic int.	Task		Schuller et al. (2009)
DD	Dyadic int.	Spontaneous		Cohn et al. (2009)
DUEL	Dyadic int.	Spontaneous, task	BM	Hough et al. (2016)
FreeTalk	Multi-party int.	Spontaneous		Scherer et al. (2009)
ILHAIRE	Combination of databases	Combination	BM	McKeown et al. (2013)
MANHOB	Single part.	Video, posed	TM	Petridis et al. (2013)
MMLI	Multi-party int.	Task	BM, resp	Niewiadomski et al. (2013)
MMI-V	Single part.	Task		Valstar and Pantic (2010)
SEMAINE	Dyadic int.	Task		McKeown et al. (2012)

Table 1: Existing databases containing laughter. From left to right: Scenarios (elicited laughter, Laughter during meetings, laughter during dyadic interactions and combinations), Modalities (BM = Body movement, resp = respiration, TM = Thermal camera, all databases in this table contained video and audio material, databases differ in how they obtain body movement information), Reference.

3.1 Main characteristics

Although the databases we described in the previous section are very valuable, we aim to make a new database available for researchers that contributes in several new ways to the already existing work. We will sum the new contributions here. We aimed to:

- Capture multimodal data on different types of laughter during dyadic interaction.
- Capture physiological data including inertial movement data, electrocardiogram (ECG) and skin conductance (SCR) signals. This will be described in greater detail in the subsection measurements.
- Capture facial expressions and body movement at a high definition setting and at a relatively high frame rate.
- Give researchers the tools to investigate links between laughter and personality by supplying data from multiple personality questionnaires.

We will now go into further details of the database.

3.2 Participants

Students of a course on Affective Computing (at the University of Twente) were asked to participate in the data collection and to find another participant from outside the course to participate in a data collection session. In total 32 participants participated over 16 sessions. From this group, 6 participants are not included in the database as they did not give consent. This results in a database of 26 participants (age m= 24, stdev = 2.3 years), consisting of 14 male and 12 female participants. Most of the participating pairs were at least to some extent familiar with each other. A diverse range of nationalities are represented in the database. The majority of the participants has the Dutch nationality (N=17) but participants with Indian, Greek, Taiwanese, Italian or German nationalities were also included. They were asked to speak English to each other during conversations. The participants were almost all university students from the University of Twente.

3.3 Measurements

To capture useful data, all participants were equipped with a similar set of microphones, cameras and sensors. Visual data of the face and upper body was captured with two Panasonic HC-V180EG-K cameras. Videos were shot with a resolution of 1920 x 1080 at 50 hz. Audio data was recorded with a Shure BLX14E-M17 wireless microphone set (in combination with a Zoom H6) to capture audio in high quality. Each participant wore a lavelier microphone around the neck.

Physiological data was captured using Shimmer sensor units.The IMU (Inertial Movement Unit) and both other units capture inertial movement with 9DoF inertial sensing and spatial understanding, which is obtained through accelerometers, a gyroscope and a magnetometer. Additionally the GSR+ unit obtains data on the Galvanic Skin Response (skin conductance response) and the EXG unit retrieves electrocardiograph data.

In addition to sensor data, questionnaire data was retrieved to get better insight in the participants personality and preferences. The TIPI (Gosling et al., 2003) and the IPIP-50-R (Goldberg et al., 2006) were collected to get more insight in the personality of participants. Both scaled have been proven to have reasonably to good psychometric qualities (Jonason et al., 2011; Holmes and Wood, 2009; Romero et al., 2012). The TAS-20 (Bagby et al., 1994a,b) was deployed to make predictions about the introspective qualities of participants. A demographic questionnaire was also presented to the participants. All questionnaires were filled in directly after participants signed the informed consent form.

3.4 Recording Set-up

After completing the consent form and questionnaires, participants were placed in a room at a table. They were placed directly opposite of each other, and sat at the side of the table that corresponded with the letters they were given (A or B). Cameras were placed on both sides of the table on fixed locations, pointing towards the participant sitting on the other side of the table. At the start of each session the cameras were manually adjusted in height and the zoom function was used to find the most optimal frame. The lavalier microphones were attached to a cord which was hung around the neck of participants. For the physiological modalities each participant was equipped with three separate Shimmer devices, the EXG+ unit, the GSR+ unit and the IMU unit. These were placed respectively, on the torso, the wrist of the dominant hand and the wrist of the non-dominant hand. See Figure 1 for an example of two participants during the session.

3.5 Recording Protocol

All participants performed several tasks during the data collection. Instructions and questionnaires for these tasks were printed in booklets. There were two slightly different versions of these booklets, to make sure that participants had different instructions during the rounds in the second task of this session. This difference in instructions was used because during the second round participants were placed in certain roles of either trying to make the other person laugh as much as possible or just being a conversation partner. In the next round the roles were reversed. Task instructions were verbally rehearsed by the experimenter and special effort was made to make sure that participants understood all instructions because some tasks were more ambiguous in nature then others. After each task, except for the first task, the participants filled in a questionnaire. The questionnaire contained questions on whether they laughed, how funny they rated themselves during that task and how funny they rated the other participant during that task.

The first task consisted of a modified ‘survival task. Participants were instructed to imagine that they were stranded on an uninhabited island and needed to construct a shared list of 10 items that they would take with them. The task ended after 3 minutes were passed. During these interactions, spontaneous laughter often occurred. The data of this task is missing in the database in three sessions.

The second task of the session contained two rounds, each round having a duration of 2 minutes. In the first round each participant had either the instructions in their booklet to start a conversation or had the instructions to make the other person laugh as much as possible. In the second round the participants switched their role. participants were blind to the other participants instructions for each round and were asked not to talk about their specific instructions.

In the third task of the session participants were instructed to tell each other jokes. This part of the session contained three identical rounds and one round with slightly adjusted instructions. Each of the three identical rounds consisted of both participants telling each other a joke in turn. Participants selected three jokes from their stack of jokes that were selected beforehand by the researchers and were distributed randomly into two stacks. A few examples of jokes that were often chosen will follow;

> “Want to hear a joke about a piece of paper? Never mind... it’s tearable.”
>
> “Why can’t a bicycle stand on its own? It’s two-tired.”
>
> “What did the grape say when it was stepped on? Nothing, it just let out a little wine.”

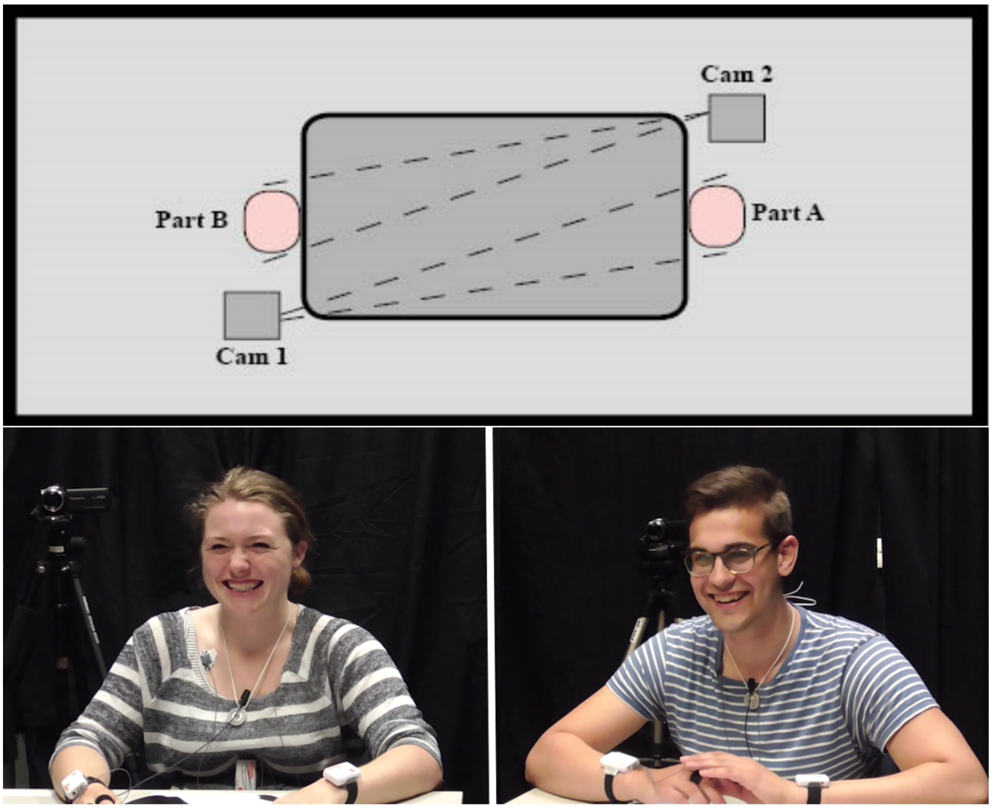


Figure 1: Example of the recording set-up. The top part of the figure shows how the cameras en participants are positioned, the lower part of the figure shows a snapshot of one of the sessions.

> "Why is it always hot in the corner of a room? Because a corner is 90 degrees."

In the last round both participants were instructed to either tell a joke they had prepared in advance, look up a joke and prepare it on the spot or select a joke that was left in their stacks.

The participants participated in a small emotion induction task involving pictures from the IAPS database (Lang et al., 1997). This task is currently not described here since it is out of the scope of this paper.

3.6 Synchronization of Data Streams

Since the audio, video and physiological data is captured with multiple devices, manual synchronization was needed. To ease this task, the participants of the session were instructed to do a synchronized clap at several intervals during each session. This is a procedure used in other database collections (Petridis et al., 2013) and ensures that several visual/sound and physiological cues are available for synchronization purposes. First the audio and visual data were synchronized by a research assistant in Adobe Premiere Pro.

This resulted in, after editing, 240 video clips (participant 1, participant 2 and combined) and 152 audio clips (participant 1 and participant 2). In addition a large amount of physiological data was retrieved during the sessions, more specifically electrocardiograph-, skin conductance response- and inertial measure data was retrieved. We plan to synchronize the physiological data in the near future.

4 Discussion and Future Work

The MULAI database allows researchers to answer a large variety of research questions. Some example research questions are the following.

- How do people express their laughter and does personality influence how people express themselves?
- Can we detect laughter in less traditional physiological data such as ECG and SCR?
- Does personality type of the participant correlate with the humor he or she expresses?

We hope that our database will help laughter researchers to explore one of these questions or answer other research questions.

5 Availability

An individual license is needed for access to the questionnaire-, video-, audio- and physiological data. Please note that only the available data from participants who explicitly gave permission for sharing the data with the research community will be shared. Please contact one of the authors for more information.

References

R. Michael Bagby, James D. A. Parker, and Graeme J. Taylor. 1994a. The Twenty-Item Item Selection Toronto and Cross-Validation Structure. *Journal of Psychosomatic research*, 38(1):23–32.

R. Michael Bagby, Graeme J. Taylor, and James D. A. Parkers. 1994b. The twenty-item Toronto Alexithymia Scale-II. Convergent, Discriminant, And Concurrent Validity. *Journal of Psychosomatic Research*, 38(1):33–40.

Jeffrey F. Cohn, Tomas Simon Kruez, Iain Matthews, Ying Yang, Minh Hoai Nguyen, Margara Tejera Padilla, Feng Zhou, and Fernando De la Torre. 2009. Detecting Depression from Facial Actions and Vocal Prosody. In *International Conference on Affective Computing and Intelligent Interaction and Workshops*, page 7.

Matthew Gervais and David Sloan Wilson. 2005. The Evolution and Functions of Laughter and Humor: A Synthetic Approach. *The Quarterly Review of Biology*, 80(4):241–277.

Lewis R. Goldberg, John A. Johnson, Herbert W. Eber, Robert Hogan, Michael C. Ashton, C. Robert Cloninger, and Harrison G. Gough. 2006. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1):84–96.

Samuel D. Gosling, Peter J. Rentfrow, and William B. Swann. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6):504–528.

John G. Holmes and Joanne V. Wood. 2009. Interpersonal situations as affordances: The example of self-esteem. *Journal of Research in Personality*, 43(2):250.

Julian Hough, Ye Tian, Laura De Ruiter, Simon Betz, Spyros Kousidis, David Schlangen, and Jonathan Ginzburg. 2016. DUEL : A Multi-lingual Multimodal Dialogue Corpus for Disfluency , Exclamations and Laughter. In *International Conference on Language Resources and Evaluation*, pages 1784–1788.

Peter K. Jonason, Emily A. Teicher, and David P. Schmitt. 2011. The TIPI's validity confirmed: Associations with sociosexuality and self-esteem. *Individual Differences Research*, 9(1):52–60.

P.J. Lang, M.M. Bradley, and B.N. Cuthbert. 1997. International Affective Picture System (IAPS): Technical Manual and Affective Ratings. Technical report.

I. Mccowan, J. Carletta, W. Kraaij, S. Ashby, S. Bourban, M Flynn, M Guillemot, T. Hain, J. Kadlec, V. Karaiskos, G. Lathoud, M. Lincoln, A. Lisowska, W. Post, D. Reidsma, and P. Wellner. 2005. The AMI Meeting Corpus. In *International Conference on Methods and Techniques in Behavioral Research.*, page 4.

G. McKeown, R. Cowie, W. Curran, W. Ruch, and E. Douglas-Cowie. 2013. ILHAIRE laughter database. In *4th International Workshop on Human Behavior Understanding(HBU) 2013*, October, page 251.

Gary McKeown, Michel Valstar, Roddy Cowie, Maja Pantic, and Marc Schröder. 2012. The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing*, 3(1):5–17.

Radoslaw Niewiadomski, Maurizio Mancini, Tobias Baur, Giovanna Varni, Harry Griffin, and Min S.H. Aung. 2013. MMLI: Multimodal multiperson corpus of laughter in interaction. In *International Workshop on Human Behavior Understanding*, pages 184–195.

Stavros Petridis, Brais Martinez, and Maja Pantic. 2013. The MAHNOB Laughter database. *Image and Vision Computing*, 31(2):186–202.

Estrella Romero, Paula Villar, J. Antonio Gómez-Fraguela, and Laura López-Romero. 2012. Measuring personality traits with ultra-short scales: A study of the Ten Item Personality Inventory (TIPI) in a Spanish sample. *Personality and Individual Differences*, 53(3):289–293.

S Scherer, Friedhelm Schwenker, N Campbell, and G Palm. 2009. Multimodal Laughter Detection in Natural Discourses. In *Human Centered Robot Systems*, pages 111–120.

Björn Schuller, Ronald Müller, Florian Eyben, Jürgen Gast, Benedikt Hörnler, Martin Wöllmer, Gerhard Rigoll, Anja Höthker, and Hitoshi Konosu. 2009. Being bored? Recognising natural interest by extensive audiovisual integration for real-life application. *Image and Vision Computing*, 27(12):1760–1774.

J Urbain, Elisabetta Bevacqua, and Thierry Dutoit. 2010. The AVLaughterCycle Database. In *Proceedings -The International Conference on Language Resources and Evaluation*, Section 11, pages 2996–3001.

Michel Valstar and M Pantic. 2010. Induced Disgust, Happiness and Surprise: an Addition to the MMI Facial Expression Database. In *proceedings -The International Conference on Language Resources and Evaluation*, pages 65–70.