# A Short Introduction to Laughter *

Stavros Petridis
Department of Computing
Imperial College London
London, UK

## 1 Production Of Laughter And Speech

The human speech production system is composed of the lungs, trachea, larynx, nasal cavity and oral cavity as shown in Fig. 1. Speech is simply a sound wave originating from the lungs. Air coming out of the lungs first passes through the trachea and then through the vocal tract. The shape of the vocal tract determines the speech sound produced. The position of the lips, tongue, velum, teeth and jaw (the articulators) define the shape of the vocal tract and therefore can influence the produced sound [65]. The nasal tract is usually closed but it can open when the velum is lowered.

A key component in the production of speech is the vocal cords, which are located at the top of the trachea, and define if a sound is voiced or unvoiced. If the vocal cords vibrate when air flows through the trachea then the produced sound is voiced and the rate of the vibration determines the fundamental frequency of the sound. In case they do not vibrate then the produced sound is unvoiced since an aperiodic noise-like sound is produced.

Laughter is produced by the same mechanism as speech but there is an important difference between them, speech is articulated but laughter is not. Bachorowski et al. [8] has shown that laughter mainly consists of central sounds and it is not articulated. This agrees with the suggestion by Ruch and Ekman [70] that articulation requires voluntary control over the vocal system which is not present during spontaneous laughter.

*This is based on Chapter 2 of the following PhD thesis [53] and the following journal publication [59]

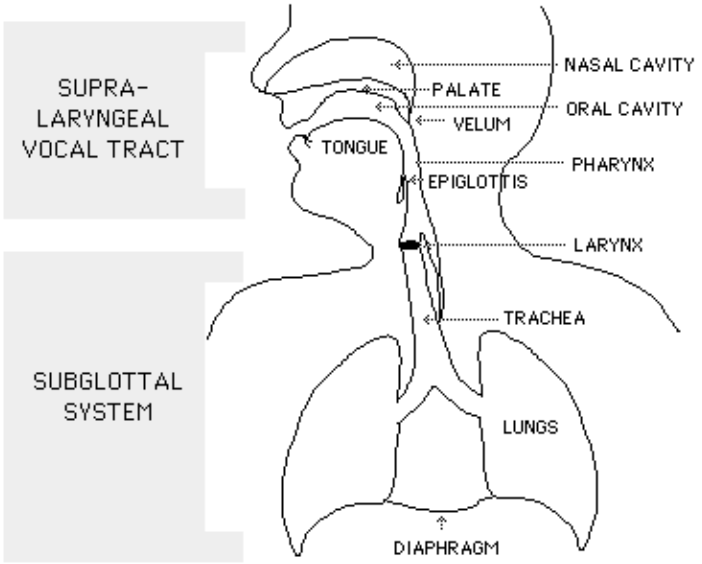


**Figure 1:** Human speech production system (from [69]).

The information in speech is carried by the audio signal but there is a strong correlation between the vocalisations and the facial configurations. The importance of the visual channel in speech perception is demonstrated by the McGurk effect [45] in which a person hearing an audio recording of /ba/ and seeing the synchronised video of another person saying /ga/ often perceives /da/. In addition, it has been shown that auditory speech becomes more audible when the involved facial movements are visible [78]. This effect clearly illustrates that speech perception is an audiovisual mechanism which is strongly influenced by the visual channel.

The same conclusion holds also for laughter. Jordan and Abedipour [32] have demonstrated that laughter is perceived as more audible when the facial expressions are visible. Participants were asked to identify laughs in an audio signal corrupted by noise and they achieved a much higher recognition accuracy when they could observe the facial expressions. However, it should be noted that this was true for negative signal-to-noise ratios, i.e., for very noisy conditions. This fact demonstrates the bimodal nature of laughter.

## 2 Research In Psychology And Linguistics

Laughter is an important social signal which plays a key role in social interactions and relationships. It is estimated that it is about 7 million years old [48], and like other social signals, it is widely believed that it evolved before speech was developed [52,70]. Several theories have been proposed trying to explain why laughter evolved, presenting contradicting views, but a common line in most of them is that laughter was an early form of communication. Owren and Bachorowski [51] suggested that laughter evolved in order to facilitate the formation and maintenance of positive and cooperative relationships in social groups. Gervais and Wilson [22] suggested that laughter was a medium for transmitting positive emotion, whereas Ramachandran [66] proposed the false alarm theory that laughter evolved in order to alert others that a potentially dangerous situation turned out to be safe. Nowadays, laughter may not be so important as speech in communicating and forming relationships with other people but it is still a very useful social signal which helps humans to express their emotions and intentions in social interactions. It is usually perceived as positive feedback, i.e., it shows joy, acceptance, agreement, but it can also be used as negative feedback, e.g., irony or even as a defence mechanism in order to deal with stressful situations, e.g., nervous laughter. Thus, laughter could still be used to strengthen or enhance relations within a group and show positive intentions towards unknown persons [23]. This is also supported by the fact that although speaking at the same time with others is generally considered rude, laughing at the same time is accepted and considered a sign of positive feedback.

The importance of laughter as a social signal is almost universally accepted and it has been confirmed by experimental studies as well. Campbell [14] presented results from the telephone conversations between Japanese speakers, showing that the speakers varied their laughing styles according to the sex and nationality of the partner. Provine [62] found that in the absence of stimulating media, e.g., television, people are about 30 times more likely to laugh, whereas they are only 4 times more likely to talk, when they are in company than when they are alone. Vettin and Todt [84] found that laughter is much more frequent in conversations than what had been previously reported in self-reported studies. They also found that the acoustic parameters do not vary randomly, but are correlated with the context in which laughter is produced, i.e. laughter in conversations vs laughter elicited by humorous situations. A similar result has been reported in [7, 8], where

the laughter rate and acoustics were shown to be associated with the sex and the familiarity with the company. In particular, male laughter seems to be highly correlated to the familiarity with his company, whereas female laughter seems to be correlated with the sex of the partner. Finally, laughter, like yawning, is highly contagious [60] and simply the sound of laughter can trigger laughter in other people.

Laughter is also used to fill pauses and regulate the flow of a conversation. Provine [61] has shown that people tend to laugh at places where punctuation would be placed in a transcript of a conversation. In a later study, it was shown that this punctuation effect is also present in deaf people while signing [63]. It has also been suggested that shared laughter is associated with topic termination [27]. On the other hand, in cases where one participant did not laugh in response to other participant's laughter but replied with further talk then the topic conversation continued whereas otherwise it might have come to an end.

Apart from a social signal, laughter is also widely believed to be beneficial for the person that produces it by influencing his/her health. Although works in the literature are rather limited there is some evidence in favour of this theory. Intense laughter leads to increased heart rate, respiratory rate and oxygen consumption like aerobic exercise does, and is followed by muscle relaxation [10]. However, these changes are temporary. It has also been found that laughter inhibits increase in the glucose levels of type 2 diabetic patients [26]. It has also been demonstrated that laughter increased pain tolerance in children [77].

Given the significance of laughter in social interactions and in health it is not surprising that there is evidence of a strong genetic basis [62]. It turns out that at least some features of laughter can be developed without the experience of hearing / seeing laughter. Babies have the ability to laugh before they can speak [70], usually laughter emerges at about 4 months after birth [75] but an instance of laughter as early as 17 days after birth has been reported [33]. In addition, children who were born both deaf and blind still have the ability to laugh [18] and the acoustic features of laughter produced by congenitally deaf and normally hearing students are similar [43]. On the other hand, it is also believed that some features of laughter are learnt. Features like the context, frequency, intensity and interpretation of laughter are highly variable in different cultures indicating that cultural norms play a role on the use of laughter as a social interaction tool [6, 22].

Contrary to popular opinion, laughter is not unique to humans [62]. It

was noted already by Darwin that chimpanzees and other great apes also produce laugh like sounds when tickled or during play [16]. Human laughter is considered to be homologous to the apes laughter [62,83] and it is believed that it was elaborated with the transition to bipedalism. This allowed more flexibility in the coordination of vocalisations and breathing and sounds were no longer tied to single breaths [62]. In other words the constraint of few syllables per breath was lifted and allowed the formation of longer sounds which is the basis for speech and laughter. This remains the main difference between human and chimpanzee laughter [62]. In other words, several laugh syllables may occur in one exhalation in human laughter, whereas chimpanzee laughter is tied to breathing and one or few laugh syllables are produced during each brief exhalation.

Since laughter has attracted interest by researchers from many disciplines, the terminology is sometimes confusing. Ruch and Ekman [70] point out that laughter is not a term used consistently nor is it precisely defined in research. Similarly, Trouvain [79] points out that terms related to laughter are either not clearly defined or they are used in different ways in different studies.

These facts illustrate the significance of laughter and explain why it is considered one of the most important universal non-verbal vocalisations. However, it is surprising that our knowledge about laughter is still incomplete and little empirical information is available [22].

## 2.1 Laughter Segmentation

Perhaps the most popular laughter segmentation approach is the one proposed by Trouvain [79], where laughter is segmented in 3 levels as shown in Fig. 2. At the low level (segmental level) there are short segments that can be described as vowel-like sounds, usually called laugh notes, bursts, pulses, or consonant-like sounds, usually called intercall intervall or interpulse pause. At the next level (syllable level), segmental sounds are joined together to form laugh syllables, called laugh events or calls. At the higher level (phrasal level), several syllables form a laugh phrase. This laugh phrase is usually called a "bout" and is defined as a sequence of laugh syllables that is produced during one exhalation [8]. Therefore an entire laugh episode consists of several "bouts" which are separated by inhalations.

It has also been suggested that when considering the facial expression which accompanies laughter then the following three phases of laughter can be identified: onset, apex, offset [70]. Onset is the pre-vocal facial part, where

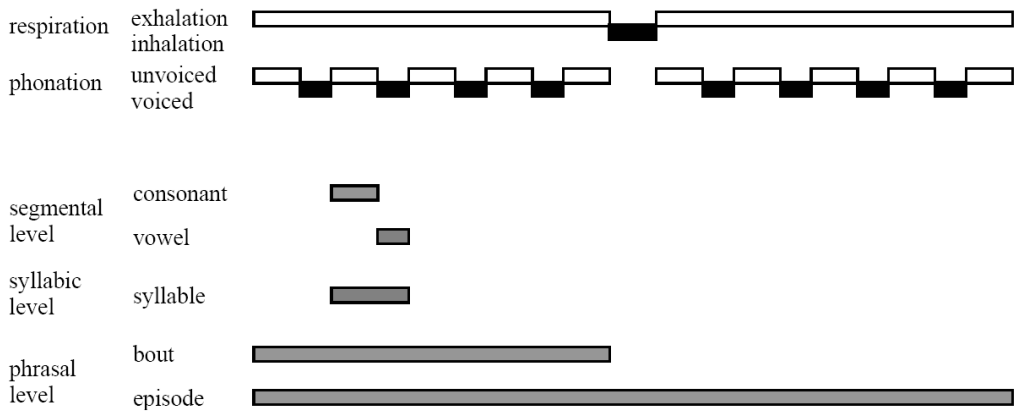


**Figure 2:** Laughter segmentation according to [79] (from [79]).

usually the mouth opens, apex is the period where the laughter vocalisation occurs, and offset is the post-vocalisation part where the mouth closes and can last for a long period. However, in practice the variability of laughter is high and it is not widely accepted when exactly laughter starts and stops [79].

## 2.2 Acoustic Properties of Laughter

Regarding the acoustics of laughter, two main streams can be distinguished in the literature. One suggests that the acoustic features of laughter are stereotyped [22,64], whereas the other suggests that its acoustics are variable and complex so laughter can be considered as a repertoire of sounds [8,36,70]. The latter theory is further supported by results in [36] where experimentally modified laughter series with varying acoustic parameters were rated closer to natural laughter than laughter series with constant parameter by listeners.

Although not all studies agree on the findings regarding acoustic parameters of laughter, the majority of them agree on some general principles. Perhaps the most studied parameter in this area is the fundamental frequency $F_0$ and almost all recent studies agree that mean $F_0$ is higher in both male and female laughter than it is in speech [8, 68, 80]. The average duration of a laughter episode varies from less than 1 second [8, 68], to approximately 2 seconds [80]. Table 1 summarises reported values in the literature regarding mean $F_0$ and laughter duration. It can be seen that the reported values for male and female mean pitch of laughter vary from 126 to 424 Hz and 160 to 535 Hz, respectively.

**Table 1:** Laughter statistics from previous works. M: Male, F: Female. - denotes no information was provided in this work. * The mean $F_0$ was reported for all subjects and there was no distinction between males and females.

| Study | No Subjects | Mean (St.Dev.) $F_0$ (Hz) | | Mean Duration (sec) |
|---|---|---|---|---|
| | | M | F | |
| Bachorowski et al. [8] | 45 M, 52 F | 284 (155) | 421 (208) | 0.87 (0.77) |
| Bickley and Hunnicutt [11] | 1 M, 1F | 138 (-) | 266 (-) | - (-) |
| Milford [46] | 15 M, 15F | 175 (-) | 160 (-) | 1.34 (-) |
| Mowrer et al. [47] | 11 M | 126 (42.7) | - (-) | 1.22 (0.44) |
| Nwokah et al. [49] | 3 F | - (-) | 365 (28) | 2.14 (-) |
| Provine and Yong [64] | 23 M, 28 F | 276 (95) | 502 (127) | - (-) |
| Rothganger et al. [68] | 20 M, 20 F | 424 (-) | 475 (125) | 0.75 (-) |
| Vettin and Todt [84] | 4 M, 6F | 171 (-) | 315 (-) | - (-) |
| Truong and van Leeuwen [80] | - | 475* (367) | | 1.80 (1.65) |
| Petridis et al. [56] | 12 M, 10 F | 400 (96) | 535 (169) | 1.80 (2.32) |

It is common to consider laughter as a series of successive elements whose parameters are not constant but changing between or even within elements [36]. Another characteristic of laughter is the alternation of voiced and unvoiced segments with the proportion of unvoiced segments being higher in laughter than in speech [80]. Finally, it has also been reported that the intensity of laughter goes down over time [70] and due to the dynamics of respiration, the duration of laugh pulses decreases with time, whereas the duration of interpulse pause increases [70].

It is not always easy to compare the findings of different studies since laughter may be defined in a different way. A typical example is how the start and end of laughter are defined, since it is not always clear when a laughter starts and particularly when it ends. For example, it is common that a laughter ends with an audible inhalation and it is not clear if this

should be part of the laughter episode or not. Another example is the existence of a short silent pause in a laughter episode resulting in either two consecutive laughs or one complex one [79]. Therefore, using different definitions can lead to different results, as in the case of [82] where the inhalation at the end of laughter was considered to be part of the laughter episode. Consequently, the average duration of laughter was found to be 3.50 seconds which almost twice as long the average duration usually reported in the literature.

## 2.3 Types Of Laughter

Several classifications have been proposed in the literature regarding different types of laughter. The most commonly accepted one is the discrimination of laughter into two types based on its acoustics: voiced and unvoiced [7, 24]. Voiced laughter is a harmonically rich, vowel-like sound with a measurable periodicity in vocal fold vibration, whereas unvoiced laughter is a noisy exhalation through nose or mouth and the vocal folds are not involved in the production of laughter. These two broad categories are characterised by significant variability. Therefore, Bachorowski et al. [8] proposed the distinction of unvoiced laughs into two classes: unvoiced snort-laughter, where the sound exits through the nose, and unvoiced grunt-like laughter, where sounds exit through the mouth. [1]. In the same study, voiced laughter is called song-like since it usually resembles the typical "ha-ha-ha" laughter consisting of multiple repeating vowel-like sounds. There are also mixed laughs which contain both voiced parts and unvoiced grunt / snort-like sounds.

Another classification has been proposed by Campbell et al. [15], which does not label an entire laughter episode but just laughter segments. They described 4 different laughter segments: voiced, chuckle, breathy and nasal, and they assumed that each laughter is composed of different combinations of these segments.

It has been demonstrated that different types of laughter have different functions in social interactions. Grammer and Eibl-Eibesfeldt [24] found that male interest was partly predicted by the number of voiced laughs produced by female partners. The opposite does not hold and this result has also been confirmed by Bachorowski and Owren [7]. The latter study also demonstrated that voiced laughter always elicited more positive evaluations than unvoiced laughter. It is also believed that voiced laughter is directly related to the experience of positive affect, whereas unvoiced laughter is used to negotiate social interactions [28]. Except judging social signals like in-

terest, the distinction between voiced and unvoiced laughter could be useful for judging the mirth of the laughter. This could be used for assessing the hilarity of observed material like movies and tagging the material in question accordingly (see [58] for a preliminary study).

Laughter is also divided into two types, spontaneous and voluntary. It has been shown that spontaneous stimulus-driven laughter and voluntary laughter involve separate neural pathways [30] similarly to spontaneous and voluntary facial expressions [21].

Regarding the types of spontaneous laughter, it has been suggested that a distinction should be made between Duchenne, which is stimulus-driven, e.g., response to humour or tickling, and emotionally valenced, and non-Duchenne which is emotionless laughter [22,34]. Stimulus for Duchenne laughter is considered any unexpected event that is perceived as non-dangerous in a social context [22] . In other words, Duchenne laughter is linked with positive emotional experience and it is stimulus driven, whereas non-Duchenne usually refers to conversation laughter in the absence of stimulus and therefore it is emotionless. It has even been suggested that non-Duchenne laughter is a learnt skill which has achieved automisation and appears to be spontaneous but in fact it is voluntary laughter [22]. However, in the vast majority of previous works [8, 36, 62, 84] this distinction has been ignored treating both types of laughter in the same way.

Another special type of laughter is speech-laughter in which speech and laughter occur at the same time. However, it has been suggested that this type has different characteristics and should be considered as a different class of non-linguistics vocalisation [38, 79]. It is also not easy to record speech laughs given that it is hard to elicit them in a lab setting and they are less frequent than laughter, for example speech-laughter is 10 times less frequent than laughter [40]. Therefore, it is not surprising that due to the lack of data research in speech-laughter is rather limited.

Finally, it has long been debated whether smile and laughter are the two extremes in the same continuum as suggested in [51]. Evidence in favour of this theory were presented in [20], where it was reported that electrical stimulation in the anterior part of the human supplementary motor area can elicit laughter. At low currents a smile was produced, while at higher currents laughter was present and its duration and intensity was dependent on the level of stimulation. Despite such evidence, this theory is not always accepted [22].

# 3 Automatic Laughter Classification / Detection

Relatively few works exist in the literature on automatic laughter classification / detection [1]. These are summarised in Tables 2 and 3. It can be seen that there is a lack of a benchmark dataset based on which different methods could be compared. The use of different datasets in combination with the use of different performance measures makes the comparison of different approaches almost impossible. Further, it can be seen from Tables 2 and 3 that both static and dynamic modelling approaches have been attempted. For dynamic modelling, Hidden Markov Models (HMM) are commonly used just as is the case in automatic speech recognition. This is mainly due to suitability of HMMs to represent temporal characteristics of the phenomenon. For static modelling, Support Vector Machines (SVM) and Neural Networks (NN) are the most commonly used tools in this field. Unlike automatic speech recognition where HMMs usually outperform static approaches, initial results on presegmented episodes using static models were very promising and that explains why these methods are still commonly used. This is also confirmed by Schuller et al. [73], who have shown that the performance of SVMs is comparable to that of HMMs for the classification of non-linguistic vocalisations. Another study [55] comparing NNs and coupled HMMs for discrimination of laughter-vs speech and posed-vs-spontaneous-smiles has come to a similar conclusion.

Regarding the audio features, several different features have been used with the most popular being the standard features used in automatic speech recognition, Mel-Frequency Cepstral Coefficients (MFCC) and Perceptual Linear Predictive (PLP) features. Pitch and energy, which have been used in emotion recognition from speech [85] are commonly used as well.

From Tables 2 and 3 it can also be seen that the vast majority of the attempts towards automatic laughter classification / detection used only audio information, i.e., visual information carried by facial expressions of the observed person is ignored. Recently, few works on audiovisual laughter detection have been reported, which use information from both the audio and visual channel (see Table 3 and the end of this section).

[1] This section is based on Chapter 2 of the following PhD thesis [53] and the following journal publication [59] and contains works up to 2012.

## 3.1 Audio-only Laughter Classification / Detection:

In this category, works can be divided into two groups. Those which focus on the detection of laughter in an unsegmented audio stream or on the discrimination between several non-linguistic vocalisations in presegmented audio episodes (where each episode contains exactly one of the target non-linguistic vocalisations), and those which perform general audio segmentation / classification into several audio categories, which are usually not non-linguistic vocalisations, e.g., music, applause, etc, and one of the classes is laughter. In the first group there are usually two approaches:

1. Laughter detection / segmentation, e.g., [35,37,41], where the aim is to segment an unsegmented audio stream into laughter and non-laughter segments.
2. Laughter-vs-speech classification / discrimination, e.g., [42,73,80], where the aim is to correctly classify presegmented episodes of laughter and speech.

One of the first works on laughter detection is that of Kennedy and Ellis [35], who trained SVMs with MFCCs, spatial cues, and modulation spectrum features (MSFs) to detect group laughter, i.e., when more than a certain percentage of participants are laughing. They used the ICSI corpus achieving true positive and false positive rates of 87% and 13% respectively. However, inconsistent results were obtained when the system was tested on unseen datasets from NIST RT-04 [5]. Truong and van Leeuwen [81] used cepstral features (PLP) for laughter segmentation in meetings. Gaussian Mixture Models (GMM) were trained for speech, laughter and silence and the system was evaluated on the ICSI corpus achieving an EER of 10.9%. Laskowski and Schultz [41] present a system for the detection of laughter and its attribution to specific participants in multi-channel recordings. Each participant can be in one of the three states (silence, speech, laughter), and the aim is to decode the vocal activity of all participants simultaneously. HMMs are used with MFCCs and energy features. The system is tested on the ICSI meeting corpus. To reduce the amount of states that a multi-party conversation can have, they apply minimum duration constraints for each vocalisation, and overlap constrains which assume that no more than a specific number of participants speak or laugh at the same time. The F1 rate achieved is 34.5%. When tested on unseen datasets, the F1 is less than 20%, but the system does not rely on manual pre-segmentation. Knox et al. [37] used MFCCs,

**Table 2:** Previous works on audio-only and audiovisual laughter classification. A: Audio, V: Video, L: Laughter, NL: Non-Laughter, S: Speech, SL: Speech-Laugh, NT: Neutral, Subj: Number of Subjects, Y: Yes, N: No, CV: Cross Validation, SI: Subject Independent, CR: Classification Rate, TP: True Positive rate, FP: False Positive Rate, EER: Equal Error Rate, R: Recall, PR: PRecision, ER: Error Rate. When no information is provided in a study then this is denoted by ?

| Study | A/V | Classifier | Features | Dataset | Size | Testing | SI | Classes | Performance |
|---|---|---|---|---|---|---|---|---|---|
| **Classification** | | | | | | | | | |
| Truong & van Leeuwen (2005) [80] | A | SVM, GMM | PLP, Pitch&Energy, Pitch&Voicing, MSF | ICSI (Bmr,Bed), CGN [50] | L: 3264, S: 3574 | Train: L:2422, S:2680 Test: L:894, S:842 | Y(Bed, CGN) | Laughter / Speech | EER: Bmr - 2.6% Bed - 2.9% CGN - 7.5% |
| Campbell (2005) [15] | A | HMM | ? | ESP [13] | L: 3000 | ? | ? | 4 Types / Segments of Laughter | CR: 81% Segments, 75% Laughter |
| Schuller et al. (2008) [73] | A | SVM, HMM, HCRF | PLP, MFCC | AVIC [74] | 2901 examples L: 261, Subj: 21 | 3-fold Stratified CV | Y | 5 classes | CR: 80.7% R: 87.7 PR: 75.1% |
| Lockerd & Mueller (2002) [42] | A | HMM | Spectral Coefficients | Own | L: 40, S: 210 Subj: 1 | Train: 70% Test: 30% | N | Laughter / Speech | CR: 88% |
| Reuderink et al. (2008) [67] | AV | A: GMMs, HMMs V: SVMs | A:RASTA-PLP, V: Shape Parameters | AMI [44] | L: 60, S: 120 Subj: 10 | 2 x 15-fold CV | N | Laughter / Speech | EER: 14.2% AUC: 0.93 |
| Batliner et al. (2009) [9] | A | SVMs | A: MFCC,Pitch Energy, ZCR | FAU AIBO [76] | L: 176 SL: 100 Subj: 51 | Leave-one-subject-out CV | Y | Laughter / Speech / Speech-Laugh | CR: 77.6% |
| Petridis & Pantic (2008) [57] | AV | NNs | A: PLP, V: Facial Points Distances | AMI | L: 40 S: 56 Subj: 8 | Leave-one-subject-out CV | Y | Laughter / Speech | R: 86.9% PR: 76.7% |
| Petridis & Pantic (2010) [54] | AV | NNs | A: MFCC, V: Shape Parameters | AMI, SAL [17] | AMI: 124 L 154 S Subj: 10 SAL: 94 L 177 S Subj: 15 | Cross-Database | Y | Laughter / Speech | F1 L: 95.4% (SAL) 76.3% (AMI) |
| Petridis & Pantic (2011) [59] | AV | NNs | A: MFCC,Pitch, Energy, ZCR V: Shape Parameters | AMI, SAL | AMI: 124 L 154 S Subj: 10 SAL: 94 L 177 S Subj: 15 | Cross-Database | Y | Laughter / Speech | F1 L: 96.6% (SAL) 72.7% (AMI) |

**Table 3:** Previous works on audio-only and audiovisual laughter detection. A: Audio, V: Video, L: Laughter, NL: Non-Laughter, S: Speech, NT: Neutral, Subj: Number of Subjects, Y: Yes, N: No, CV: Cross Validation, SI: Subject Independent, CR: Classification Rate, TP: True Positive rate, FP: False Positive Rate, EER: Equal Error Rate, R: Recall, PR: PRecision, ER: Error Rate. When no information is provided in a study then this is denoted by ?

| Study | A/V | Classifier | Features | Dataset | Size | Testing | SI | Classes | Performance |
|---|---|---|---|---|---|---|---|---|---|
| **Detection / Segmentation** | | | | | | | | | |
| Kennedy & Ellis (2004) [35] | A | SVM | MFCC, Mod. Spectrum, Spatial Cues | ICSI(Bmr) [31], NIST RT-04 [5] | L: 1926 (ICSI) 44 (NIST), Subj: 8 | CV (ICSI) Train: 26 meetings Test: 3 meetings | Y (NIST) | Laughter / Non-Laughter | ICSI TP: 87% FP: 13% |
| Truong & van Leeuwen (2007) [81] | A | GMMs | PLP | ICSI(Bmr) | L: 91min, S: 93min Subj: 10 | Train: 26 meetings Test: 3 meetings | N | Laughter / Speech / Silence | EER: 10.9% |
| Laskowski & Schultz (2008) [41] | A | HMM | MFCC, Energy | ICSI (Bmr, Bro, Bed) | NT: 716.2min S: 94.4min L: 16.6min Subj: 23 | Train: 26 meetings Bmr Test: 3 meetings Bmr, Bro, Bed | N | Laughter / Speech / Neutral | F1: 34.5% |
| Knox et al. (2008) [37] | A | NNs | MFCC, Pitch, Energy, Phones, Prosodics, MSF | ICSI (Bmr) | L: 6641 sec NL: 98848 sec | Train: 26 meetings Test: 3 meetings | N | Laughter / Non-Laughter | EER: 5.4% |
| Ito et al. (2005) [29] | AV | A: GMMs, V: LDFs | A: MFCC, V: Lip angles, lengths, Cheek mean intensities | Own | 3 dialogues, 4 - 8 min each Subj: 3 | 5-fold CV | N | Laughter / Non-Laughter | R: 71% PR: 74% |
| Escalera et al. (2009) [19] | AV | SSL | A: Pitch, Spectral Entropy V: Mouth Movements | New York Times [2] | 9 videos, 4 min each Subj: 18 | 10-fold CV | ? | Laughter / Non-Laughter | CR: 77% |
| Scherer et al. (2009) [71] | AV | ESN | A: Mod. Spectrum V: Head/Body Movements | FreeTalk [3] | 3 videos, 90 min each Subj: 4 | 10-fold CV | N | Laughter / Non-Laughter | CR: 91.5% |

pitch, energy, phones, prosodics and MSFs with neural networks in order to segment laughter by classifying audio frames as laughter or non-laughter. A window of 1 010 ms (101 frames) was used as input to the neural network and the output was the label of the centre audio frame (10 ms). The ICSI corpus was used and an equal error rate of 5.4% was achieved.

The most extensive study in laughter-vs-speech discrimination was made by Truong and van Leeuwen [80], who compared the performance of different audio-frame-level features (PLP, Pitch and Energy) and utterance-level features (Pitch and Voicing, Modulation Spectrum) using SVMs and GMMs. They used the ICSI corpus [31] and CGN corpus [50] achieving an equal error rate of 2.6% and 7.5% in subject-dependent and subject-independent experiments, respectively. Campbell et al. [15] first divided laughter into 4 classes: hearty, amused, satirical, and social, and decomposed each laughter into 4 laughter segments: voiced, chuckle, breathy, and nasal. They used HMMs to recognise these 4 laughter segments and the 4 classes of entire laugh episodes from the ESP corpus [13] resulting in classification rates of 81% and 75% respectively. Schuller et al. [73] used the AudioVisual Interest Corpus (AVIC) [74] to classify 5 types of non-linguistic vocalisations: laughter, breathing, hesitation, consent, and other vocalisations including speech. They used HMMs and Hidden Conditional Random Fields (HCRF) with PLP, MFCC and energy features, and SVMs with several statistical features, e.g., mean, standard deviation, etc., which describe the variation over time of other low level descriptors, e.g., pitch, energy, zero-crossing rate, etc. Using a 3-fold stratified cross validation they reported an overall classification rate of 80.7%. From the confusion matrix provided in [73], the recall and precision of laughter can be computed which are 87.7% and 75.1%, respectively. Lockerd and Mueller [42] used spectral coefficients and HMMs with the aim to detect when the operator of a video camera laughs. The system was trained using data of a single subject achieving a classification rate of 88%.

To the best of our knowledge, there are only two works which try to recognise different types of laughter. The first one is by Laskowski [39] who developed an HMM system based on his previous work [41] to detect silence, speech, unvoiced laughter and voiced laughter in conversational interactions. His main conclusion was that modelling only voiced laughter leads to better performance than modelling all laughter. The ICSI meeting corpus was used with its standard partition for training, validation and test sets. The F1 rate for voiced laughter led to a relative improvement over the F1 rate for all laughter from 6% to 22%. The other work is by Batliner et al. [9] who used

the FAU Aibo Emotion Corpus [76], which contains children communicating with Sony's pet robot Aibo, for laughter classification. Five types of laughter were annotated, weak speech-laugh, strong speech-laugh, voiced laughter, unvoiced laughter, and voiced-unvoiced laughter. All episodes were presegmented and the goal was to discriminate between those 5 types and speech. The same methodology as in [72, 73] was used, i.e. statistics of low-level features, like MFCC, pitch, etc, were computed over the entire episode. SVMs were used for classification in a leave-one-subject out manner, resulting in a classification rate of 58.3%. Another experiment was conducted where the 2 types of speech-laugh and 3 types of laughter were merged into broader classes, speech-laugh and laughter. In this 3-class scenario the performance was significantly improved, as expected, to 77.6%.

In the second group of approaches, there are usually several classes which correspond to different sounds, e.g., laughter, applause, music, etc. Because of the nature of this problem the features used are more diverse. That includes zero crossing rate (ZCR), brightness (BRT), bandwidth (BW), Total Spectrum Power (TSP) and SubBand Powers (SBP), and Short Time Energy (STE) in addition to the standard features mentioned above. Similar classification methods as above, like SVMs [25] and HMMs [12] have been used. Since these works are not focused on laughter detection / classification, they are not described in this study in further detail.

## 3.2 Audiovisual Laughter Classification / Detection:

To the best of our knowledge there are only three works on audiovisual laughter detection and one on laughter-vs-speech discrimination and, as a consequence, the approaches followed are less diverse. The first study on audiovisual laughter detection was conducted by Ito et al. [29], who built an image-based laughter detector based on geometric features (lip lengths and angles), mean intensities in the cheek areas (grayscale images were used), and an audio-based laughter detector based on MFCC features. Linear discriminant functions (LDFs) and GMMs were used for the image-based and audio-based detectors, respectively, and the output of the two detectors were combined with an AND operator to yield the final classification for an input sample. They attained 71% recall rate and 74% precision rate using 3 sequences of 3 subjects in a person-dependent way. In a more recent work, Scherer et al. [71] used the FreeTalk corpus [3,4] to detect laughter in a meeting scenario. Due to the nature of data simple visual features were extracted

describing the face and body movement. Modulation spectrum features were extracted from the audio modality and Echo State Networks (ESN) were used as classifiers. One ESN was trained for each modality, audio and video, and the outputs were fused using a weighted sum. Using three 90-minute recordings in a 10-fold cross-validation experiment, the audiovisual approach resulted in a small absolute improvement of 1.5% in classification rate over the audio-only approach, from 89.5% to 91%. Finally, Escalera et al. [19] performed audiovisual laughter recognition from 9 dyadic video sequences from the New York Times video library [2]. The mouth of each participant was first localised in the video and then features related to the mouth movement were extracted. For audio, the commonly used fundamental frequency was used together with features derived from the spectrogram, accumulated power and spectral entropy. Detection is performed per frame using a stacked sequential learning (SSL) schema, and evaluation is performed using a 10-fold cross-validation. However, it is not clear how beneficial the audiovisual approach is since the classification rate goes down to 77% from 81% when audio-only is used. On the other hand, the recall rate for the audiovisual approach goes up to 65% from 61% when audio-only is used.

Reuderink et al. [67] used visual features based on principal components analysis (PCA) and RASTA-PLP features for audio processing for a laughter-vs-speech discrimination problem. GMMs and HMMs were used for the audio classifier, whereas SVMs were used for the video classifier. The outputs of the classifiers were fused on decision level, by weighted combination of the audio and video modalities. The system was tested in a subject-dependent way on 60 episodes of laughter and 120 episodes of speech from the AMI corpus. The audiovisual approach led to a small increase in the AUC (Area Under the ROC curve) over the best unimodal approach, which is video in this case, from 0.916 to 0.93. On the other hand, the audiovisual approach was worse than video-only in terms of the equal error rate, 14.2% and 13.3% for the audiovisual and video-only approaches, respectively.

# References

[1] `http://www.psy.vanderbilt.edu/faculty/bachorowski/laugh.htm`.

[2] `http://video.nytimes.com/`.

[3] http://freetalk-db.sspnet.eu/files/.

[4] http://www.speech-data.jp/corpora.html.

[5] NIST (2004). rich transcription 2004 spring meeting recognition evaluation, documentation. http://www.nist.gov/speech/tests/rt/rt2004/spring/.

[6] M. Apte. *Humor and laughter: An anthropological approach*. Cornell university press, 1985.

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

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

[9] A. Batliner, S. Steidl, F. Eyben, and B. Schuller. On laughter and speech laugh, based on observations of child-robot interaction. *The Phonetics of Laughing*, 2010.

[10] M. Bennett and C. Lengacher. Humor and laughter may influence health: III. laughter and health outcomes. *Evidence Based Complementary and Alternative Medicine*, 5(1):37–40, 2008.

[11] C. Bickley and S. Hunnicutt. Acoustic analysis of laughter. In *In Proc. Int'l Conf. on Spoken Language Processing*, 1992.

[12] R. Cai, L. Lu, H.-J. Zhang, and L.-H. Cai. Highlight sound effects detection in audio stream. In *Int'l Conf. on Multimedia and Expo*, volume 3, pages 37–40, 2003.

[13] N. Campbell. Recording techniques for capturing natural everyday speech. In *Proc. Language Resources and Evaluation Conf.*, 2002.

[14] N. Campbell. Whom we laugh with affects how we laugh. In *Workshop On the Phonetics Of Laughter*, pages 61–65, 2007.

[15] N. Campbell, H. Kashioka, and R. Ohara. No laughing matter. In *Europ. Conf. on Speech Comm. and Technology*, pages 465–468, 2005.

[16] C. Darwin. *The expression of emotions in animals and man*. 1873.

[17] E. Douglas-Cowie, R. Cowie, C. Cox, N. Amir, and D. Heylen. The Sensitive Artificial Listener: an induction technique for generating emotionally coloured conversation. In *Workshop on Corpora for Research on Emotion and Affect*, pages 1–4, 2008.

[18] I. Eibl-Eibesfeldt. The expressive behavior of the deaf-and-blind born. *Social Communication and Movement: Studies of Interaction and Expression in Man and Chimpanzee*, pages 163–193, 1973.

[19] S. Escalera, E. Puertas, P. Radeva, and O. Pujol. Multi-modal laughter recognition in video conversations. In *IEEE CVPR Workshops 2009*, pages 110–115. IEEE.

[20] I. Fried, C. Wilson, K. MacDonald, and E. Behnke. Electric current stimulates laughter. *Nature*, 391(6668):650–650, 1998.

[21] M. Gazzaniga and C. Smylie. Hemispheric mechanisms controlling voluntary and spontaneous facial expressions. *Journal of Cognitive Neuroscience*, 2(3):239–245, 1990.

[22] M. Gervais and D. Wilson. The evolution and functions of laughter and humor: a synthetic approach. *The Quarterly Review of Biology*, 80(4):395, 2005.

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

[24] K. Grammer and I. Eibl-Eibesfeldt. The ritualisation of laughter. *Die Natürlichkeit der Sprache und der Kultur. Bochum: Brockmeyer*, pages 192–214, 1990.

[25] G. Guo and S. Li. Content-based audio classification and retrieval by support vector machines. *IEEE Trans. on Neural Networks*, 14(1):209–215, 2003.

[26] K. Hayashi, T. Hayashi, S. Iwanaga, K. Kawai, H. Ishii, S. Shoji, and K. Murakami. Laughter lowered the increase in postprandial blood glucose. *Diabetes care*, 26(5):1651–1652, 2003.

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

[28] W. Hudenko, W. Stone, and J. Bachorowski. Laughter differs in children with autism: An acoustic analysis of laughs produced by children with and without the disorder. *Journal of Autism and Developmental Disorders*, 39(10):1392–1400, 2009.

[29] A. Ito, W. Xinyue, M. Suzuki, and S. Makino. Smile and laughter recognition using speech processing and face recognition from conversation video. In *Intern. Conf. on Cyberworlds, 2005*, pages 8–15, 2005.

[30] M. Iwase, Y. Ouchi, H. Okada, C. Yokoyama, S. Nobezawa, E. Yoshikawa, H. Tsukada, M. Takeda, K. Yamashita, M. Takeda, et al. Neural substrates of human facial expression of pleasant emotion induced by comic films: a pet study. *Neuroimage*, 17(2):758–768, 2002.

[31] A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, et al. The ICSI meeting corpus. In *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing*, volume 1, pages 364 – 367, 2003.

[32] T. Jordan and L. Abedipour. The importance of laughing in your face: Influences of visual laughter on auditory laughter perception. *Perception*, 39(9):1283–1285, 2010.

[33] K. Kawakami, K. Takai-Kawakami, M. Tomonaga, J. Suzuki, T. Kusaka, and T. Okai. Origins of smile and laughter: A preliminary study. *Early Human Development*, 82(1):61–66, 2006.

[34] D. Keltner and G. Bonanno. A study of laughter and dissociation: Distinct correlates of laughter and smiling during bereavement. *Journal of Personality and Social Psychology*, 73(4):687, 1997.

[35] L. Kennedy and D. Ellis. Laughter detection in meetings. In *NIST Meeting Recognition Workshop*, 2004.

[36] S. Kipper and D. Todt. The role of rhythm and pitch in the evaluation of human laughter. *J. Nonverb. Behavior*, 27(4):255–272, 2003.

[37] M. Knox, N. Morgan, and N. Mirghafori. Getting the last laugh: Automatic laughter segmentation in meetings. In *Proc. of INTERSPEECH*, pages 797–800, 2008.

[38] K. J. Kohler. 'speech-smile', 'speech-laugh', 'laughter' and their sequencing in dialogic interaction. *Phonetica*, 465(1-2):1–18, 2008.

[39] K. Laskowski. Contrasting emotion-bearing laughter types in multiparticipant vocal activity detection for meetings. In *Proc. IEEE ICASSP*, pages 4765–4768, 2009.

[40] K. Laskowski and S. Burger. Analysis of the occurrence of laughter in meetings. In *Proc. INTERSPEECH*, pages 1258–1261, 2007.

[41] K. Laskowski and T. Schultz. Detection of laughter-in-interaction in multichannel close-talk microphone recordings of meetings. *Lecture Notes in Computer Science*, 5237:149–160, 2008.

[42] A. Lockerd and F. Mueller. Lafcam: Leveraging affective feedback camcorder. In *CHI, Human factors in computing systems*, pages 574–575, 2002.

[43] M. Makagon, E. Funayama, and M. Owren. An acoustic analysis of laughter produced by congenitally deaf and normally hearing college students. *The Journal of the Acoustical Society of America*, 124:472483, 2008.

[44] I. McCowan, J. Carletta, W. Kraaij, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, and V. Karaiskos. The AMI meeting corpus. In *Int'l. Conf. on Methods and Techniques in Behavioral Research*, pages 137–140, 2005.

[45] H. McGurk and J. Macdonald. Hearing lips and seeing voices. *Nature*, 264(5588):746–748, 1976.

[46] P. Milford. *Perception of laughter and its acoustical properties.* PhD thesis, Pennsylvania State University., 1980.

[47] D. Mowrer, L. LaPointe, and J. Case. Analysis of five acoustic correlates of laughter. *Journal of Nonverbal Behavior*, 11(3):191–199, 1987.

[48] C. Niemitz. Visuelle zeichen, sprache und gehirn in der evolution des menscheneine entgegnung auf mcfarland (visual signs, language and the brain in the evolution of humansa reply to mcfarland. *Z. Sem.*, 12:323–336, 1990.

[49] E. Nwokah, H. Hsu, P. Davies, and A. Fogel. The integration of laughter and speech in vocal communication: A dynamic systems perspective. *Journal of Speech, Language, and Hearing Research*, 42(4):880, 1999.

[50] N. Oostdijk. The Spoken Dutch Corpus: Overview and first evaluation. In *Proc. Int'l. Conf. Language Resources and Evaluation*, pages 887–894, 2000.

[51] M. J. Owren and J.-A. Bachorowski. The evolution of emotional expression: A selfish-gene account of smiling and laughter in early hominids and humans. In *Emotion: Current issues and future directions*, pages 152–191, New York: Guilford, 2001. Mayne, T. J. and Bonanno, G. A. (Eds.).

[52] A. Pentland. *Honest signals: how they shape our world*. The MIT Press, 2008.

[53] S. Petridis. *Audiovisual Discrimination Between Laughter and Speech*. PhD thesis, Imperial College London, 2012.

[54] S. Petridis, A. Asghar, and M. Pantic. Classifying laughter and speech using audio-visual feature prediction. In *Proc. IEEE ICASSP*, pages 5254–5257, 2010.

[55] S. Petridis, H. Gunes, S. Kaltwang, and M. Pantic. Static vs. dynamic modeling of human nonverbal behavior from multiple cues and modalities. In *Proc. ICMI*, pages 23–30. ACM, 2009.

[56] S. Petridis, B. Martinez, and M. Pantic. The mahnob laughter database. *Image and Vision Computing Journal*, 31(2):186–202, February 2013.

[57] S. Petridis and M. Pantic. Audiovisual discrimination between laughter and speech. In *IEEE Int'l Conf. Acoustics, Speech, Signal Processing*, pages 5117–5120, 2008.

[58] S. Petridis and M. Pantic. Is this joke really funny? Judging the mirth by audiovisual laughter analysis. In *Proc. IEEE Intl Conf. Multimedia & Expo*, pages 1444–1447, 2009.

[59] S. Petridis and M. Pantic. Audiovisual discrimination between speech and laughter: Why and when visual information might help. *IEEE Trans. on Multimedia*, 13(2):216 –234, April 2011.

[60] R. Provine. Contagious laughter: Laughter is a sufficient stimulus for laughs and smiles. *Bulletin of the Psychonomic Society*, 1992.

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

[62] R. Provine. *Laughter: A Scientific Investigation*. New York: Viking, 2000.

[63] R. Provine and K. Emmorey. Laughter Among Deaf Signers. *Journal of Deaf Studies and Deaf Education*, 11(4):403, 2006.

[64] R. Provine and Y. Yong. Laughter: a stereotyped human vocalization. *Ethology*, 89(2):115–124, 1991.

[65] L. Rabiner and R. Schafer. *Digital processing of speech signals*. Prentice Hall, 1978.

[66] V. Ramachandran. The neurology and evolution of humor, laughter, and smiling: the false alarm theory. *Medical hypotheses*, 51(4):351–354, 1998.

[67] B. Reuderink, M. Poel, K. Truong, R. Poppe, and M. Pantic. Decision-level fusion for audio-visual laughter detection. *Lecture Notes in Computer Science*, 5237:137 – 148, 2008.

[68] H. Rothgänger, G. Hauser, A. Cappellini, and A. Guidotti. Analysis of laughter and speech sounds in Italian and German students. *Naturwissenschaften*, 85(8):394–402, 1998.

[69] P. Rubin and E. Vatikiotis-Batkson. Measuring and modeling speech production. *Animal acoustic communication: sound analysis and research methods*, pages 251–290, 1998.

[70] W. Ruch and P. Ekman. The expressive pattern of laughter. In *Emotions, Qualia, and Consciousness*, pages 426–443, 2001.

[71] S. Scherer, F. Schwenker, N. Campbell, and G. Palm. Multimodal laughter detection in natural discourses. *Human Centered Robot Systems*, pages 111–120, 2009.

[72] B. Schuller, A. Batliner, D. Seppi, S. Steidl, T. Vogt, J. Wagner, L. Devillers, L. Vidrascu, N. Amir, L. Kessous, and V. Aharonson. The relevance of feature type for the automatic classification of emotional user states: Low level descriptors and functionals. In *INTERSPEECH*, pages 2253–2256, 2007.

[73] B. Schuller, F. Eyben, and G. Rigoll. Static and dynamic modelling for the recognition of non-verbal vocalisations in conversational speech. *Lecture Notes in Computer Science*, 5078:99–110, 2008.

[74] B. Schuller, R. Mueller, B. Hoernler, A. Hoethker, H. Konosu, and G. Rigoll. Audiovisual recognition of spontaneous interest within conversations. In *Proc. ACM Int'l Conf. Multimodal Interfaces*, pages 30–37, 2007.

[75] L. Srofe and E. Waters. The ontogenesis of smiling and laughter: A perspective on the organization of development in infancy. *Psychological Review*, 83(3):173, 1976.

[76] S. Steidl. *Automatic classification of emotion-related user states in spontaneous children's speech.* Logos-Verl., 2009.

[77] M. Stuber, S. Hilber, L. Mintzer, M. Castaneda, D. Glover, and L. Zeltzer. Laughter, humor and pain perception in children: a pilot study. *Evidence-based Complementary and Alternative Medicine*, 97:1–6, 2007.

[78] W. Sumby and I. Pollack. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26(2):212–215, 1954.

[79] J. Trouvain. Segmenting phonetic units in laughter. In *Proc. Int'l Conf. Phonetic Sciences*, pages 2793–2796, 2003.

[80] K. P. Truong and D. A. van Leeuwen. Automatic discrimination between laughter and speech. *Speech Communication*, 49(2):144–158, 2007.

[81] K. P. Truong and D. A. Van Leeuwen. Evaluating laughter segmentation in meetings with acoustic and acoustic-phonetic features. In *Workshop on the Phonetics of Laughter*, 2007.

[82] J. Urbain, E. Bevacqua, T. Dutoit, A. Moinet, R. Niewiadomski, C. Pelachaud, B. Picart, J. Tilmanne, and J. Wagner. The AVLaughterCycle database. In *Proceedings of the International Conference on Language Resources and Evaluation*, 2010.

[83] J. van Hooff and R. A. Hinde. A comparative approach to the phylogeny of laughter and smiling. *Non-verbal communication*, 1972.

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

[85] Z. Zeng, M. Pantic, G. Roisman, and T. Huang. A survey of affect recognition methods: Audio, visual and spontaneous expressions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009.