Introducing AmuS: The Amused Speech Database

Kevin El Haddad^{1(✉)}, Ilaria Torre², Emer Gilmartin³, Hüseyin Çakmak¹,
Stéphane Dupont¹, Thierry Dutoit¹, and Nick Campbell³

¹ University of Mons, Mons, Belgium

{kevin.elhaddad,huseyin.cakmak,stephane.dupont,
thierry.dutoit}@umons.ac.be

² Plymouth University, Plymouth, UK

ilaria.torre@plymouth.ac.uk

³ Trinity College Dublin, Dublin, Ireland

{gilmare,nick}@tcd.ie

Abstract. In this paper we present the **AmuS** database of about three hours worth of data related to amused speech recorded from two males and one female subjects and contains data in two languages French and English. We review previous work on smiled speech and speech-laughs. We describe acoustic analysis on part of our database, and a perception test comparing speech-laughs with smiled and neutral speech. We show the efficiency of the data in **AmuS** for synthesis of amused speech by training HMM-based models for neutral and smiled speech for each voice and comparing them using an on-line CMOS test.

Keywords: Corpora and language resources · Amused speech · Laugh · Smile · Speech synthesis · Speech processing · HMM · Affective computing · Machine learning

1 Introduction

Recognition and synthesis of emotion or affective states are core goals of affective computing. Much research in these areas deals with several emotional or affective states as members of a category of human expressions – grouping diverse states such as anger, happiness and stress together. The emotive states are either considered as discrete classes [27,33] or as continuous values in a multidimensional space [1,31], or as dynamically changing processes over time. Such approaches are understandable and legitimate as the goal in most cases is to build a single system that can deal with several of the emotional expressions displayed by users.

However, such approaches often require the same features to be extracted from all classes for modeling purposes. Emotional states vary greatly in their expression, and in how they manifest in different subjects, posing difficulties in providing uniform feature sets for modeling a range of emotions – for example, happiness can be expressed with laughs, which can be a periodic expressions

with a certain rhythm, while disgust is usually expressed with continuous non-periodic expressions. Some studies focused on a single emotion such as stress [18] and amusement [12], with the aim to build a model of one emotional state. In this study, we explore the expression of amusement through the audio modality.

Limiting models to audio cues has several advantages. Speech and vocalization are fundamental forms of human communication. Audio is easier and less computationally costly to collect, process, and store than other modalities such as video or motion capture. Many applications rely on audio alone to collect user spoken and affective information. Conversational agents in telephony or other ‘hands/eyes free’ platforms use audio to “understand” and interact with the user. Many state-of-the-art robots such as NAO cannot display facial expressions, and rely on audio features to express affective states.

Amusement is very frequently present in human interaction, and therefore data are easy to collect. As amusement is a positive emotion, collection is not as complicated as the collection of more negative emotions such as fear or disgust, for ethical reasons. In addition, the ability to recognize amusement can be very useful in monitoring user satisfaction or positive mood.

Amusement is often expressed through smiling and laughter, common elements of our daily conversations which should be included in human-agent interaction systems to increase naturalness. Laughter accounts for a significant proportion of conversation – an estimated 9.5% of total spoken time in meetings [30]. Smiling is very frequent, to the extent that smiles have been omitted from comparison studies as they were so much more prevalent than other expressions [6].

Laughter and smiling have different social functions. Laughter is more likely to occur in company than in solitude [19], and punctuates rather than interrupts speech [37]; it frequently occurs when a conversation topic is ending [3], and can show affiliation with a speaker [20], while smiling is often used to express politeness [22]. In amused speech, both smiling and laughter can occur together or independently. As with laughter, smiling can be discerned in the voice when co-occurring with speech [8, 40]. The phenomenon of laughing while speaking is sometimes referred to as speech-laugh [43], smiling while speaking has been called smiling voice [36, 42], speech-smiles [25] or smiled speech [14]. As listeners can discriminate amused speech based on the speech signal alone [29, 42], the perception of amused speech must be directly linked to these components and thus to parameters which influence them, such as duration and intensity. McKeown and Curran showed an association between laughter intensity level and humor perception [32]. This suggests that the intensity level of laughter and, by extension, of all other amused speech components may be a particularly interesting parameter to consider in amused speech.

In this paper we present the **AmuS** database, intended for use as a resource for the analysis and synthesis of amused speech, and also for purposes such as amusement intensity estimation. **AmuS** is publicly and freely available for research purposes, and can be obtained from the first author¹. It contains several

¹ **AmuS** is available at: <http://tcts.fpms.ac.be/~elhaddad/AmuS/>.

amused speech components from different speakers and in different languages (English and French) in sufficient quantity for corpus-based statistically robust studies. The database contains recorded sentences of actors producing smiled speech along with corresponding neutrally pronounced speech (i.e., with no specific emotion expressed) as well as laughter and speech-laugh from some of these speakers adequate for analysis and synthesis. The difference from previous work on the topic is, to the best of our knowledge, the quantity and nature of material provided.

We also describe the following experiments, which demonstrate the efficiency of the **AmuS** database for different problems:

- Acoustic study of smiled speech data from **AmuS** and comparison of results with previous findings.
- On-line perception tests, comparing speech-laugh, smiled speech and neutral sentences in terms of the arousal intensity scale.
- Evaluation results of HMM-based amused speech synthesis systems trained with the **AmuS** data.

Below we briefly review the characteristics of smiled speech and speech-laugh, outline the recording protocols we employed and describe the database of recordings. Finally, we present and discuss results of the acoustic study carried out on the recordings and results of the perception and evaluation experiments.

2 Motivations and Contributions

For the purposes of this study, we classify amused speech into two categories, smiled speech and speech-laugh, although finer distinctions are possible.

Smiled Speech. As an emotional expression, when it co-occurs with speech, smiling is formed not only by labial spreading but also by additional modifications of the vocal tract. Lasarczyk and Trouvain [29] conducted a perception test asking participants to rate the “smiliness” of vowels synthesized using an articulatory synthesizer. These were synthesized by modifying three parameters: lip spreading, fundamental frequency and the larynx length. The vowels synthesized by modifying all three parameters were perceived as most “smiley”. In previous work, amused smiled speech was perceived as more amused than synthesized neutral or spread-lips speech, and the synthesized spread-lips sentences were also perceived as more amused than the synthesized neutral sentences [12]. This suggests that simply spreading the lips can give an impression of amusement. In earlier studies, Tartter found that recorded spread-lips sentences were indeed perceived as “smiled” [40]. Thus, although several factors contribute to emotional smiled speech, spreading the lips while speaking can play a strong role in its expression.

Smiled speech has also been studied at the acoustic level, with emphasis on analysis of the fundamental frequency (f_0) and the first three formants (F1, F2, F3). f_0 was found to be higher in smiled than neutral vowels in several studies [2, 16, 28, 40]. Emond et al. and Drahota et al. also compared mean f_0 , f_0

height and f_0 range but did not find any systematic change between non-smiled and smiled recorded speech [8, 14]. In these two studies, they compared whole sentences rather than isolated vowels or short words. However, after perception tests using these same sentences were performed, both works reported a certain correlation or relationship between f_0 increase and perception of smiling.

Several acoustic studies show that the modification of the vocal tract caused by smiling during speech affects formant frequencies [8, 12, 16, 29, 38, 40, 42], although this effect varies somewhat from one study to another. Several authors [2, 8, 16, 38, 40, 41] report higher F2 in smiled than non-smiled speech. In addition, in work dealing with smiled speech in which the speaker was asked not to express any specific positive emotion while smiling was not reported to be subject to any spread-lips stimuli, authors reported a less important or even absent increase of F1 [12, 16, 38, 40]. Barthel and Quené [2] used naturalistic data of dyadic conversations, in the form of spontaneous smiles by speakers, to acoustically compare smiled and neutral utterances. They report increases in the first three formants but only obtained significant results for F2 and for rounded vowels. In [42], a comparison was made between smiled and non-smiled data in a dataset of naturally occurring conversation. An increase in F3 was reported from neutral to smiled speech data, as well as an average increase of formants for rounded vowels. Drahota et al. [8] also used conversational data for their study of different kinds of smile, where they compared the distance between mean formant values for smiled and neutral speech, e.g. F2-F1 for smiled speech compared versus F2-F1 neutral speech, instead of differences in formant height for smiled speech and neutral speech. They report that the more times a speaker is perceived as not smiling, the larger the difference F3-F2. In a previous study, we found variation, in the form of increase and decrease in both F1 and F2 values, for amused smiled speech with a greater effect on F2 [12].

The work cited above leads to several conclusions. f_0 is reported as higher in smiled than neutral speech or to be an important parameter of smiled speech perception. Lip spreading predominantly affects F2 and has no or little effect on F1. Relationships between formants (such as their distance) should be considered as potentially important parameters for discriminating smiled from neutral speech. Finally, vocal tract deformation in smiled speech varies in different ways depending on context and speaker.

Smile detection or recognition systems often form part of a multi-class emotion classification system (smile would then be associated to positive emotions such as happiness) [24], and are also found in facial detection systems rather than systems based on audio features [23]. Recognition has been largely based on feature extraction for later classification [15]. The latest works on recognition using audio cues rely on feature learning systems using the power of deep learning to learn an internal representation of raw data with respect to a given task [17].

Smiled speech synthesis is generally reported in the literature as part of a voice adaptation or voice conversion system aiming to generate different emotions [46]. Among the very few studies on smiled speech synthesis is the articulatory synthesis approach of Lasarcyk and Trouvain [29], which only considers smiled (but not specifically amused) vowels.

The scarcity of work related to amused or even smiled speech is probably due to the lack of relevant data and the complexity of collecting it. Several databases containing emotional speech exist, such as [4, 5]. However, to the best of our knowledge, these databases contain smiled speech related to happiness or joy rather than amusement, and thus might not be representative of amused smiled speech. For analysis and tasks like voice conversion it is very useful to have utterances expressed in a certain emotion (in this case amused smiled speech) and as neutral utterances; especially if the pairs came from the same speaker. Such data can be found in [4] (again for happy and not precisely amused smiled speech), but the amount for a single speaker might not be sufficient for current systems data requirements.

We have collected a database of amused smiled speech and corresponding neutral speech sentences from different speakers (male and female) in English and French. This database contains enough data for analysis and for use in machine learning-based systems. For a deeper understanding of amused smiled speech we also recorded corresponding spread-lips data.

Speech-Laughs. Speech-laughs are instances of co-occurrence of laughter and speech. The phenomenon has not been clearly defined in the literature. Provine reports laughter occurring mostly at the extremities of a sentences [37], while Nowkah et al. [34] report that up to 50% of conversational laughter is produced simultaneously with speech. Kohler notes the occurrence of speech - smiled speech - speech-laugh - laughter and vice versa sequencing in amusing situations in a small-scale study and mentions the need for further investigations [25]. Trouvain proposes an acoustic account of speech-laughs and describes types of speech-laughs found in the analyzed data. He mentions the intervention of breath bursts during speech and notes the presence of “tremor” in voiced segments. He also investigates whether speech-laughs and smiled speech can be placed on a continuum of amused expression, while commenting that the variety of functions possible for laughter and the individuality of laughter among subjects would lead to significant variation in the forms of speech-laugh encountered [43]. This introduces another important aspect to be considered in relation to amused speech: the continuity between smiled speech and laughter. Are these nothing more than different levels of amusement expressions? Can intensity levels of amusement be mapped to combinations of amused speech components detected or not in an utterance? No clear answer has been given concerning this aspect of amused speech, although Trouvain rejected the hypothesis of a smile-laughter continuum [43]. Dumpala et al. found that f_0 was higher in laughter than in speech-laughs and higher in speech-laughs than in neutral speech [9].

There has been very little work on synthesis and recognition of speech-laughs and smiled speech. Dumpala et al. [9] present a speech-laugh/laughter discrimination system. Oh and Wang [35] tried real-time modulation of neutral speech to make it closer to speech-laughs, based on the variation of characteristics such as pitch, rhythm and tempo. However, no evaluation of the naturalness of that approach has been reported.

The **AmuS** database contains amused smiled speech, different types of speech-laughs and also laughs. We hope that the collection will thus reflect

the fact that laughter can interrupt or intermingle with speech [25, 34, 43], or happen at the extremities of sentences [37]. Several laughter databases can be found related to isolated laughter [10], but we know of no speech-laugh database suitable for machine learning-based work or analysis. **AmuS** will facilitate research into the use of these components in amused speech and their relation with amused speech arousal/intensity, and also into the smile-laughter continuum.

3 Database

The **AmuS** database contains recordings of acted smiled speech and corresponding neutral utterances from three different speakers in two different languages, English and French. For some speakers, spread-lips speech was also recorded. It also contains, for some speakers, speech-laughs and laughter data. The speech-laughs were semi-acted since they were collected during the smiled speech recordings but without the speakers being explicitly asked to utter them. The laughs were spontaneously expressed since they were elicited with appropriate stimuli. The data were recorded in quiet rooms and resampled to 16 kHz for uniformity. A more detailed description is given below and the data are summarized in Table 1.

Smiled Speech: To provide comparative neutral and smiled speech, speakers were asked to read the same sentences neutrally (not expressing any particular emotion) and while sounding amused but without laughing. Two readers were also asked to read the same utterances while spreading their lips without trying to sound happy or amused. This was done to obtain a set of the same sentences read in different speech styles for comparison. Noisy (saturated, or containing artifacts) or wrongly pronounced data were removed from the dataset. The final dataset is shown in the “Speech Styles” columns of Table 1. Voices from two males (M) and one female (F) were collected. SpkA and SpkB are French native speakers while SpkC is a British English native speaker. The sentences used to record SpkA were phonetically balanced. The other French data were recorded using a subset of these sentences. For English, a subset of the CMU Arctic Speech Database [26] was used. All utterances were then force-aligned with their phonetic transcriptions using the HMM ToolKit (HTK) software [44]. The transcriptions were stored in label files in the HTK label format. SpkC’s annotations were also manually checked. The amount of data available, to the best of our knowledge, is comparable (SpkB and SpkC) and in some cases superior (SpkA) to currently available databases.

Speech-Laughs: To represent the effect of laughter altering speech and thus creating speech-laughs, we collected different types of speech-laughs from SpkA and SpkC. The speech-laughs were collected from the smiled speech recordings where actors produced them intuitively when trying to sound amused. In total 161 speech-laughs were collected from SpkA and SpkC. These were divided into

two groups based on how they were produced. The first group contains tremor-like sounds happening in vowels only. These were previously investigated by us and will be referred to as chuckling (or shaking) vowels [21]. The second type contains bursts of air appearing usually at the end of a syllable or between a consonant and a vowel. A total of 109 and 52 instances were collected from the first and second type respectively.

Laughs were collected from SpkB. Since these laughs are to be used in sentences, we needed them to occur during an utterance. So, SpkB was asked to sustain a vowel while watching funny videos. He eventually laughed while pronouncing a vowel. A total of 148 laughs were recorded. These types of laughs proved to be efficient to produce synthesized amused speech [13]. The perceived arousal level or affective intensity of these laughs were then annotated in an on-line experiment. Annotators were presented with 35 randomly picked laughs and asked to rate how amused the laughs sounded, on a scale from 0 to 4 (0 begin not amused and 4 being very amused). The annotators were free to logout at any time. A total of 22 annotators took part in this experiment and each laugh was annotated on average 5.08 times.

Table 1. AmuS database content description. The numbers represent the number of utterance collected. M=Male, F=Female, Lang=Language, SL=Speech-laugh, L=laughs, * = these are the same data since the laughs came from the same speaker.

Speakers	Speech styles			Lang	SL	L
	N	Sm	Sp			
SpkA (M)	1085	1085	-	Fr	48	-
SpkB-Fr (M)	249	199	199	Fr	-	148*
SpkB-Eng (M)	180	213	-	Eng	-	148*
SpkC (F)	170	84	152	Eng	113	-

4 Acoustic Analysis

In this section, we present data analysis on the acoustic effects of smiling on speech, comparing neutral speech vowels to amused smiled and spread-lips speech vowels. We considered 15 and 16 vowels from French and English respectively.

The f_0 and first three formants (F_1 , F_2 and F_3) were calculated for each sentence in each of the three speech styles using a sliding Hamming window of 25 ms length, shifted by 10 ms using the Snack library [39]. F_2-F_1 and F_3-F_2 were also calculated for each vowel. Pairs were formed of the same sentences from amused smiled and neutral speech. Since the durations of two corresponding sentences in a pair were different, Dynamic Time Warping (DTW) was applied at the phoneme level to align each of the extracted parameters. Values for neutral speech were subtracted from those for amused smiled speech. The mean values of all the differences obtained were then calculated. The same method was then applied between the spread-lips and neutral speech styles. Table 2 shows the results of these mean differences. These are expressed in percentage of variation

with respect to the neutral speech values (e.g. 5% represents an increase of 5% of the neutral speech value, -5% a decrease). This table shows the mean percentage obtained as well as the mean standard deviation.

Table 2. Mean difference of the acoustic parameters extracted between the neutral and the other two styles. These are expressed in % of variation with respect to the neutral speech, e.g. 56% $\pm$ 20% shows an increase of 56% of and from the neutral speech values on average with 20% standard deviation.

Speakers	Amused smiled - neutral				Spread-lips - neutral	
	SpkA	SpkB-Fr	SpkB-Eng	SpkC	SpkB-Fr	SpkC
F_0	56% $\pm$ 20%	8% $\pm$ 16%	48% $\pm$ 21%	16% $\pm$ 18%	4% $\pm$ 15%	14% $\pm$ 13%
F1	13% $\pm$ 13%	-6% $\pm$ 14%	-2% $\pm$ 15%	2% $\pm$ 10%	3% $\pm$ 10%	3% $\pm$ 14%
F2	-1% $\pm$ 9%	-2% $\pm$ 12%	-5% $\pm$ 9%	0.6% $\pm$ 7%	2% $\pm$ 6%	11% $\pm$ 11%
F3	-0.2% $\pm$ 10%	-5% $\pm$ 9%	-2% $\pm$ 9%	-1% $\pm$ 5%	-2% $\pm$ 6%	1% $\pm$ 8%
F2 - F1	-4% $\pm$ 15%	7% $\pm$ 17%	-2% $\pm$ 15%	1% $\pm$ 11%	3% $\pm$ 11%	16% $\pm$ 16%
F3 - F2	11% $\pm$ 26%	-4% $\pm$ 23%	13% $\pm$ 27%	-3% $\pm$ 20%	-6% $\pm$ 20%	-7% $\pm$ 23%

A 95% confidence interval paired Student's t-test was used to study the statistical significance of the results in Table 2. The mean value was calculated for each of the acoustic parameters and for all the vowels for each sentence. The set of values obtained for amused smiled and for the spread-lips styles were each compared to the neutral style for each speaker. All proved to be significantly different except for the spread-lips vs neutral F3 values of SpkB-Fr.

As can be seen from Table 2, f_0 increased in all cases, congruently with previous studies. Instead, no common pattern could be noticed in the formants even for the same speaker in two different languages (SpkB-Fr and SpkB-Eng), although F3 seems to be decreasing in all cases. Regarding the spread lips, instead, the pattern observed is more consistent between speakers (all changing in the same way except for F3). Thus, amused smiles affect speech in different ways with different vocal tract modifications. Since these are read sentences, further perceptual studies should be made to compare the naturalness and amusement perceived. This might help to understand the acoustic variations better.

5 Perception Test

A perception test was carried out in order to compare the level of perceived amusement from neutral (N), smiled speech (Sm) and speech-laugh (Sl). For this, we used sentences from SpkA and SpkC. 16 sentences were randomly selected from SpkA, of which 6 were the same sentences for all three speech styles, while the remaining 10 were chosen randomly. From SpkC, 15 sentences were randomly picked from each speech style, since unfortunately no identical sentences could be found in the smiled and speech-laugh styles. The sentences were then paired so that each style could be compared to the other two (N vs. Sm, Sm vs. Sl and N vs. Sl). Except for the 6 same sentences from SpkA, all the others were randomly paired.

These sentences were then presented to 20 on-line raters as a Comparative Mean Opinion Score (CMOS) test. Each rater was given thirty pairs of sentences chosen randomly and asked to rate which one sounded more amused. The raters were given seven possible choices each time: three on the right in favor of the utterance on the right, three on the left in favor of the utterance on the left and one in the middle, representing neutrality (i.e., both audio utterances sound the same). Each choice was mapped to an integer ranging from -3 to $+3$. The scores were as follows: 0.130 in favor of Sl when compared to Sm, 0.92 for Sm when compared to N and 1.135 in favor of Sl when compared to N.

Thus, Sl sentences were perceived as more amused than both Sm and N, and Sm more amused than N. Although the scores obtained in favor of Sl when compared to Sm are not very high, this result suggests that containing the tremor and/or air outburst in speech is more likely to increase the amusement intensity level perception.

6 Synthesis Evaluation

The smiled and neutral speech from all voices were also used to train Hidden Markov Model (HMM)-based speech synthesis systems [45] to generate a smiled amused voice and a neutral voice for each speaker. The systems obtained from SpkA and SpkB-Fr were previously trained and evaluated in [11] and [12], respectively. The systems from SpkC and SpkB-Eng were trained and evaluated for the purpose of this study.

HMMs were trained for each of the smiled and neutral speech styles and for each speaker and language separately. Their topology was identical in each case and consisted of 5 states left-to-right HMMs with no skip. Gaussian Mixture Models (GMM) were used to model the observation probabilities for each state. These were single multivariate Gaussian distributions with diagonal covariance matrices since a unique voice was being modeled in each case. The features used for training were the Mel Generalized Cepstral Coefficients (order 35, $\alpha = 0.35$ and $\gamma = 0$) and the f_0 , along with their derivatives and double derivatives. The features were extracted using a 25 ms wide window shifted by 5 ms, using the Snack library. Except for SpkA, for which two HMM models were created independently for each of the smiled and neutral voices, all the other models were built using adaptation via the CMLLR algorithm [7], since SpkB-Fr, SpkB-Eng and SpkC have fewer samples. For adaptation, a large dataset was used as a source on which to adapt a smaller dataset, the target. Thus, for the French voice, the neutral data of SpkA were used to build the source model and the SpkB-Fr neutral and smiled targets. For the English voice, the RMS and SLT voices from the CMU Arctic Speech Database were used as source data respectively to target SpkB-Eng and SpkC (for both neutral and smiled). The implementation was done using the HTS toolkit [45].

Synthesized smiled and neutral sentences were used for a comparison study on an amusement level scale for each of the three voices. The same sentences were synthesized in the neutral and smiling conditions for each speaker and paired for a CMOS on-line evaluation. Participants were asked the same question as in the

CMOS of Sect. 5, and in all cases the synthesized smiled sentences generated were perceived as more amused than the neutral ones, indicating that this database is suitable to train a parametric speech synthesizer such as an HMM-based one (0.24 in favor of the synthesized amused smiled sentences).

7 Conclusion

In this article we presented an amused speech database containing different amused speech components which can be grouped into the broad categories of smiling and laughter. We also reviewed a state-of-the-art for the acoustic studies relative to smiled speech and previous literature concerning speech-laugh. Our database was also used for acoustic studies and well as synthesis and perceptual evaluations. Perception tests suggest that the presence of “low level” laughs occurring in amused speech, known as speech-laugh, may increase the perception of amusement of the utterance. In the future we plan on improving this database by adding naturalistic non-acted amusement expressions, which could be used for research purposes such as affective voice conversion or speech synthesis.

References

1. Ressel, J.A.: A circumplex model of affect. *J. Pers. Soc. Psychol.* **39**, 1161 (1980)
2. Barthel, H., Quené, H.: Acoustic-phonetic properties of smiling revised-measurements on a natural video corpus. In: *Proceedings of the 18th International Congress of Phonetic Sciences* (2015)
3. Bonin, F., Campbell, N., Vogel, C.: Time for laughter. *Knowl.-Based Syst.* **71**, 15–24 (2014)
4. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W.F., Weiss, B.: A database of German emotional speech. In: *Interspeech*, vol. 5, pp. 1517–1520 (2005)
5. Busso, C., Bulut, M., Lee, C.C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J., Lee, S., Narayanan, S.S.: IEMOCAP: interactive emotional dyadic motion capture database. *J. Lang. Res. Eval.* **42**(4), 335–359 (2008)
6. Chovil, N.: Discourse oriented facial displays in conversation. *Res. Lang. Soc. Interact.* **25**(1–4), 163–194 (1991)
7. Digalakis, V.V., Rtschev, D., Neumeyer, L.G.: Speaker adaptation using constrained estimation of Gaussian mixtures. *IEEE Trans. Speech Audio Process.* **3**(5), 357–366 (1995)
8. Drahota, A., Costall, A., Reddy, V.: The vocal communication of different kinds of smile. *Speech Commun.* **50**(4), 278–287 (2008)
9. Dumpala, S., Sridaran, K., Gangashetty, S., Yegnanarayana, B.: Analysis of laughter and speech-laugh signals using excitation source information. In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 975–979, May 2014
10. Dupont, S., et al.: Laughter research: a review of the ILHAIRE project. In: Esposito, A., Jain, L.C. (eds.) *Toward Robotic Socially Believable Behaving Systems - Volume I. ISRL*, vol. 105, pp. 147–181. Springer, Cham (2016). doi:[10.1007/978-3-319-31056-5_9](https://doi.org/10.1007/978-3-319-31056-5_9)

11. El Haddad, K., Cakmak, H., Dupont, S., Dutoit, T.: An HMM approach for synthesizing amused speech with a controllable intensity of smile. In: IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Abu Dhabi, UAE, 7–10 December 2015
12. El Haddad, K., Dupont, S., d'Alessandro, N., Dutoit, T.: An HMM-based speech-smile synthesis system: an approach for amusement synthesis. In: International Workshop on Emotion Representation, Analysis and Synthesis in Continuous Time and Space (EmoSPACE), Ljubljana, Slovenia, 4–8 May 2015
13. El Haddad, K., Dupont, S., Urbain, J., Dutoit, T.: Speech-laugh: an HMM-based approach for amused speech synthesis. In: International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015
14. Émond, C., Ménard, L., Laforest, M.: Perceived prosodic correlates of smiled speech in spontaneous data. In: Bimbot, F., Cerisara, C., Fougeron, C., Gravier, G., Lamel, L., Pellegrino, F., Perrier, P. (eds.) INTERSPEECH, pp. 1380–1383. ISCA (2013)
15. Eyben, F., Scherer, K., Schuller, B., Sundberg, J., André, E., Busso, C., Devillers, L., Epps, J., Laukka, P., Narayanan, S., Truong, K.: The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. *IEEE Trans. Affect. Comput.* **7**(2), 190–202 (2015). Open access
16. Fagel, S.: Effects of smiling on articulation: lips, larynx and acoustics. In: Esposito, A., Campbell, N., Vogel, C., Hussain, A., Nijholt, A. (eds.) Development of Multimodal Interfaces: Active Listening and Synchrony. LNCS, vol. 5967, pp. 294–303. Springer, Heidelberg (2010). doi:[10.1007/978-3-642-12397-9_25](https://doi.org/10.1007/978-3-642-12397-9_25)
17. Fayek, H.M., Lech, M., Cavedon, L.: Evaluating deep learning architectures for speech emotion recognition. *Neural Netw.* **92**, 60–68 (2017)
18. Garcia-Ceja, E., Osmani, V., Mayora, O.: Automatic stress detection in working environments from smartphones' accelerometer data: a first step. *IEEE J. Biomed. Health Inform.* **20**(4), 1053–1060 (2016)
19. Glenn, P.: *Laughter in Interaction*, vol. 18. Cambridge University Press, Cambridge (2003)
20. Haakana, M.: Laughter and smiling: notes on co-occurrences. *J. Pragmat.* **42**(6), 1499–1512 (2010)
21. Haddad, K.E., Çakmak, H., Dupont, S., Dutoit, T.: Amused speech components analysis and classification: towards an amusement arousal level assessment system. *Comput. Electr. Eng.* (2017). <http://www.sciencedirect.com/science/article/pii/S0045790617317135>
22. Hoque, M., Morency, L.-P., Picard, R.W.: Are you friendly or just polite? – analysis of smiles in spontaneous face-to-face interactions. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACHI 2011. LNCS, vol. 6974, pp. 135–144. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-24600-5_17](https://doi.org/10.1007/978-3-642-24600-5_17)
23. Ito, A., Wang, X., Suzuki, M., Makino, S.: Smile and laughter recognition using speech processing and face recognition from conversation video. In: 2005 International Conference on Cyberworlds (CW 2005), pp. 437–444, November 2005
24. Kim, Y., Provost, E.M.: Emotion spotting: discovering regions of evidence in audiovisual emotion expressions. In: Proceedings of the 18th ACM International Conference on Multimodal Interaction, ICMI 2016, New York, NY, USA, pp. 92–99. ACM (2016)
25. Kohler, K.J.: “Speech-smile”, “speech-laugh”, “laughter” and their sequencing in dialogic interaction. *Phonetica* **65**(1–2), 1–18 (2008)
26. Kominek, J., Black, A.W.: The CMU arctic speech databases. In: Fifth ISCA Workshop on Speech Synthesis (2004)

27. Koolagudi, S.G., Rao, K.S.: Emotion recognition from speech: a review. *Int. J. Speech Technol.* **15**(2), 99–117 (2012)
28. Kraut, R.E., Johnston, R.E.: Social and emotional messages of smiling: an ethological approach. *J. Pers. Soc. Psychol.* **37**(9), 1539 (1979)
29. Lasarczyk, E., Trouvain, J.: Spread lips+ raised larynx+ higher f0= Smiled Speech?- an articulatory synthesis approach. In: *Proceedings of ISSP* (2008)
30. Laskowski, K., Burger, S.: Analysis of the occurrence of laughter in meetings. In: *Proceedings of the 8th Annual Conference of the International Speech Communication Association (Interspeech 2007)*, Antwerp, Belgium, pp. 1258–1261, 27–31 August 2007
31. Bradley, M.M., Greenwald, M.K., Petry, M.C., Lang, P.J.: Remembering pictures: pleasure and arousal in memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **18**, 379 (1992)
32. McKeown, G., Curran, W.: The relationship between laughter intensity and perceived humour. In: *Proceedings of the 4th Interdisciplinary Workshop on Laughter and Other Non-verbal Vocalisations in Speech*, pp. 27–29 (2015)
33. Ming, H., Huang, D., Xie, L., Wu, J., Dong, M., Li, H.: Deep bidirectional LSTM modeling of timbre and prosody for emotional voice conversion. In: *17th Annual Conference of the International Speech Communication Association, Interspeech 2016*, 8–12 September 2016, San Francisco, CA, USA, pp. 2453–2457 (2016)
34. Nwokah, E.E., Hsu, H.C., Davies, P., Fogel, A.: The integration of laughter and speech in vocal communicationa dynamic systems perspective. *J. Speech Lang. Hear. Res.* **42**(4), 880–894 (1999)
35. Oh, J., Wang, G.: Laughter modulation: from speech to speech-laugh. In: *INTER-SPEECH*, pp. 754–755 (2013)
36. Pickering, L., Corduas, M., Eisterhold, J., Seifried, B., Eggleston, A., Attardo, S.: Prosodic markers of saliency in humorous narratives. *Discourse process.* **46**(6), 517–540 (2009)
37. Provine, R.R.: Laughter punctuates speech: linguistic, social and gender contexts of laughter. *Ethology* **95**(4), 291–298 (1993)
38. Robson, J., Janet, B.: Hearing smiles-perceptual, acoustic and production aspects of labial spreading. In: *XIVth Proceedings of the XIVth International Congress of Phonetic Sciences*, vol. 1, pp. 219–222. *International Congress of Phonetic Sciences* (1999)
39. Sjölander, K.: The Snack Sound Toolkit [computer program webpage] (consulted on September, 2014). <http://www.speech.kth.se/snack/>
40. Tartter, V.: Happy talk: perceptual and acoustic effects of smiling on speech. *Percept. Psychophys.* **27**(1), 24–27 (1980)
41. Tartter, V.C., Braun, D.: Hearing smiles and frowns in normal and whisper registers. *J. Acoust. Soc. Am.* **96**(4), 2101–2107 (1994)
42. Torre, I.: Production and perception of smiling voice. In: *Proceedings of the First Postgraduate and Academic Researchers in Linguistics at York (PARLAY 2013)*, pp. 100–117 (2014)
43. Trouvain, J.: Phonetic aspects of “speech laughs”. In: *Oralité et Gestualité: Actes du colloque ORAGE, Aix-en-Provence. L’Harmattan, Paris*, pp. 634–639 (2001)
44. Young, S.J., Young, S.: The HTK hidden Markov model toolkit: design and philosophy. In: *Entropic Cambridge Research Laboratory, Ltd.* (1994)
45. Zen, H., Nose, T., Yamagishi, J., Sako, S., Masuko, T., Black, A., Tokuda, K.: The HMM-based speech synthesis system (HTS) version 2.0. In: *Proceeding 6th ISCA Workshop on Speech Synthesis (SSW-6)*, August 2007
46. Zen, H., Tokuda, K., Black, A.W.: Statistical parametric speech synthesis. *Speech Commun.* **51**(11), 1039–1064 (2009)