INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Use of Vowels in Discriminating Speech-laugh from Laughter and Neutral Speech

Sri Harsha Dumpala, P. Gangamohan, Suryakanth V. Gangashetty and B. Yegnanarayana

International Institute of Information Technology, Hyderabad, India
sriharsha.dumpala@research.iiit.ac.in, gangamohan.p@students.iiit.ac.in
svg@iiit.ac.in, yegna@iiit.ac.in

Abstract

In natural conversations, significant part of laughter co-occurs with speech which is referred to as speech-laugh. Hence, speech-laugh will have characteristics of both laughter and neutral speech. But it is not clearly evident how acoustic properties of neutral speech are influenced by its co-occurring laughter. The objective of this study is to analyze the acoustic variations between vowel regions of laughter, speech-laugh and neutral speech. The features based on excitation source characteristics extracted at epochs are considered in this study. Features extracted in the vowel regions of speech-laugh exhibit deviations from that of laughter and neutral speech. These deviations in feature values are exploited to discriminate speech-laugh from laughter and neutral speech. Two different datasets consisting of conversational speech and meeting recordings are used in this analysis. Experimental results show that the discrimination between the three classes obtained by considering vowel regions is better than that of considering the complete utterance.

Index Terms: Speech-laugh, laughter, vowels, epochs, excitation source.

1. Introduction

Current state-of-the-art automatic speech recognition (ASR) systems perform fairly well for clearly articulated speech. But their performance degrades for spontaneous speech. This can be attributed not only to the large intra-speaker and inter-speaker variations, but also to the occurrence of non-verbal events (like laughter, cry, cough etc.) and their speech co-occurring counterparts. Hence analyzing such events is essential in developing sophisticated ASR systems. Laughter and its speech co-occurring counterpart, i.e., speech-laugh, forms one such pair which occurs most frequently in natural conversations.

Laughter, being a highly variable non-verbal event, often co-occurs with speech resulting in segments referred to as speech-laugh [1]. Speech-laugh is not just laughter superimposed on articulation, but it is formed as a result of complex vocal production exhibiting characteristics of both laughter and neutral speech [2]. This makes analysis of speech-laugh segments highly difficult. But the frequent occurrence of these segments emphasize the need for their analysis. These segments occur so frequently that more than 50 percent of the laughter segments in natural conversations are speech-laughs [2]. Analysis of these segments not only helps in developing sophisticated ASR systems, but also useful in applications such as emotion detection and emotive speech synthesis.

In recent times, much emphasis was laid on analysis and detection of laughter segments in continuous speech [3-9]. Laughter segments were analyzed at bout, call, segment and syllable levels [3-4]. Analysis of the prosodic differences between initiating and responding laugh was reported in [5]. Laughter segments in continuous speech were detected using excitation source and spectral features [6-7]. Apart from conversational speech, analysis and detection of laughter events in multi-speaker environment was also performed [8-9]. All these works helped in developing a better intuition towards the production mechanism of laughter. But most of these works either discarded speech-laugh segments or considered them as laughter. A few studies were performed to explore the acoustic properties of speech-laugh and its variations from laughter and neutral speech [1-2, 10-14]. Mother child interactions were analyzed to find the simultaneous production of laughter and articulation in speech-laugh segments [2]. Phonetic characteristics of speech-laugh were analyzed to show the presence of reinforced respiratory activity in speech-laugh segments [10].

Differences between laughter and speech-laugh were analyzed using both excitation source, and vocal tract system based features [1, 11-13]. Acoustic features such as formant frequencies, voice quality, fundamental frequency (F_0) and strength of excitation were used to analyze and discriminate laughter and speech-laugh [1, 11-13]. Also, features such as pitch, intensity and rhythm were used as an initial attempt to synthesize speech-laugh from neutral speech [14]. Speech-laugh, being a speech-synchronous form of laughter, exhibits similarities and differences from both laughter and neutral speech. The main objective of this study is to analyze the discriminability of speech-laugh from laughter and neutral speech using excitation source based features extracted in vowel regions.

The organization of the paper is as follows. Section 2 explains the datasets used for analysis. Analysis of the proposed features along with the approach and the method for feature extraction is explained in Section 3. Experiments and their results are discussed in Section 4. Summary and conclusions are given in Section 5.

2. Dataset

Two databases collected in different scenarios are used to analyze the speech-laugh segments with respect to laughter and neutral speech. The datasets are explained below.

Conversational speech data is collected in English language from 14 subjects (10 male and 4 female). This data was collected at Speech and Vision lab. of IIIT-Hyderabad. This database is named as “SVL speech-laugh database”. The data was recorded by asking two speakers, who knew each other, from the group to discuss on a funny topic which helped in eliciting laughter and speech-laugh naturally. Each speaker was asked to repeat the speech-laugh utterances spoken by him in

Copyright © 2016 ISCA

http://dx.doi.org/10.21437/Interspeech.2016-1114

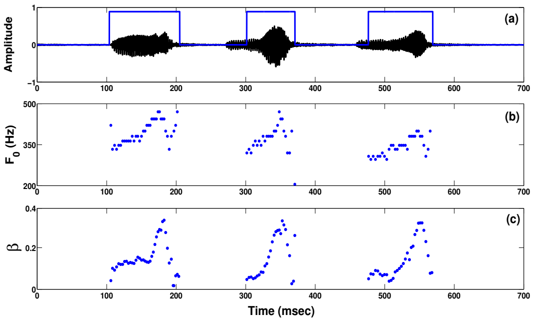

Figure 1: (a) A laughter segment with voiced/nonvoiced decision, (b) F_0 contour, (c) β contour.

his neutral speech. The data was recorded using a high quality zoom recorder, at a sampling frequency of 48 kHz in a recording room (clean environment). Each conversation was manually segmented into laughter, speech-laugh and neutral speech utterances. Overlapped speech segments were discarded for analysis purpose. The speech-laugh utterances were subjectively evaluated by 10 listeners. The subjects were asked to rate the speech-laugh utterances based on their perception between 1 and 5, where 5 refers to best and 1 refers to worst. 70 speech-laugh utterances which were rated above 4 are used for this analysis. The data also consists of 60 laughter segments.

The second database used in this analysis is AMI meeting corpus [15]. AMI meeting corpus is a multi-modal dataset consisting of 100 hours of meeting recordings. Four speakers participate in each meeting, where they discuss spontaneously among themselves on a given topic. All meetings are in English. But a large portion of the speakers are non-native English speakers, providing high degree of variability in speech pattern. Audio was collected using individual lapel microphones and headset condenser microphones. In this analysis, the headset condenser microphone data collected from each speaker is used. A dataset of 15 meetings (consisting of 25 male and 15 female speakers) recorded at Edinburgh university is considered. This dataset consists of 90 speech-laugh utterances, 100 neutral speech utterances and 65 laughter utterances. Laughter, speech-laugh and neutral speech segments uttered by each individual are collected separately for ease of analysis.

All utterances of both datasets, are manually labeled at phone level to analyze the proposed features in vowels regions.

3. Feature extraction and analysis

3.1. Method for feature extraction

Modified zero frequency filtering (MZFF) method is used to extract the excitation source based features and the epochs from the speech signal [7]. Instantaneous fundamental frequency (F_0) and strength of excitation at epochs (β) are the features used to represent the excitation source information. The rapid variations in F_0 of laughter and speech-laugh can be captured using MZFF method [7, 11]. Following is the brief description of the steps involved in the MZFF method to extract proposed features [8]. (a) Pass the speech signal through the zero frequency resonator (ZFR) with window length of 3 msec for trend removal. (b) Slope of the filtered signal calculated at the positive zero crossings gives the β values. (c) Based on the mean of β over a window length of 10 msec, the signal is divided into voiced and nonvoiced regions. (d) Each voiced region is separately passed through a ZFR. The trend removal is performed using a window length equal to the average pitch period of that voiced region, which is obtained using autocorrelation function. The resultant signal is called the ZFF signal. (e) The positive zero crossings of the ZFF signal give the epoch locations.

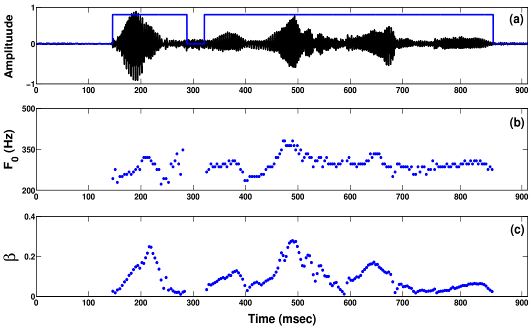

Figure 2: (a) Speech-laugh utterance, “Its really funny”, with voiced/nonvoiced decision, (b) F_0 contour, (c) β contour.

The following features are extracted from the ZFF signal around epochs. (1) F_0 is obtained by taking the reciprocal of the pitch period (T_0), where T_0 is the distance between two successive epochs. (2) β corresponds to the rate of glottal closure and is obtained by computing the slope of the ZFF signal at epochs. [16]. (3) Ratio of strength of excitation and pitch period (γ) is used as an approximate measure of the opening phase of the vocal folds [7]. γ values are obtained at every epoch by computing the ratio of β and T_0 (i.e., β/T_0) at each epoch. The features F_0, β and γ represent the excitation source information and are previously used for detection of laughter, and in discrimination of speech-laugh from laughter [7, 11]. In this work, analysis of laughter, speech-laugh and neutral speech is performed by extracting these three features at epochs. The F_0 and β contours obtained at epochs for laughter, speech-laugh and neutral speech are shown in Figs. 1, 2 and 3, respectively.

3.2. Approach for analysis of features

Terms used and the approach followed for analysis in this paper is explained below.
(a) The terms laughter (L), neutral speech (N) and speech-laugh (S) refer to pure laugh segments without any speech, plain speech without laughter and segments in which speech co-occurs with laughter, respectively.
(b) For a given utterance (either laughter, speech-laugh or neutral speech), only voiced regions, as obtained from MZFF method, are considered for analysis.
(c) For speech-laugh and neutral speech, two cases are considered.

- Complete utterance (C): In this case, the features are extracted at epochs in all voiced regions.
- Vowels (V): In this case, the features are extracted at epochs occurring only in vowel regions. The vowel regions are selected using the manual phone labels.

(d) For laughter, the features are extracted at epochs in all voiced regions. As laughter is generally of the format “haha”, “hihi”, etc., all voiced regions in laughter are also assumed as vowels [1]. Hence, two separate cases are not considered.

3.3. Analysis of features

Table 1 gives the mean (μ) and standard deviation (σ) of the feature values of laughter for both datasets. Tables 2-3 give the μ and σ of the feature values obtained for speech-laugh and

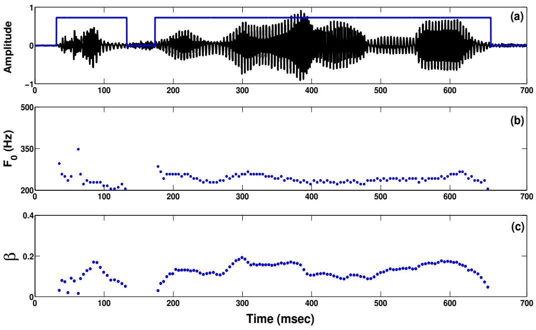

Figure 3: (a) Neutral speech utterance, "its really funny", with voiced/nonvoiced decision, (b) F_0 contour, (c) β contour.

Table 1: Mean (μ) and standard deviation (σ) of feature values of laughter for AMI and SVL speech-laugh datasets.

		AMI		SVL speech-laugh	
	Feature	μ	σ	μ	σ
Female	F$_0$	340.06	135.22	360.34	148.61
	β	129.72	81.62	121.44	85.69
	γ	70.12	58.51	69.25	56.73
Male	F$_0$	270.64	131.39	283.68	123.50
	β	74.28	58.39	73.61	53.08
	γ	29.03	25.14	26.68	20.81

neutral speech on AMI and SVL speech-laugh datasets, respectively. In Tables 2-3, S_C, S_V, N_C and N_V refer to C of speech-laugh, V of speech-laugh, C of neutral speech and V of neutral speech, respectively. The following observations can be made from Tables 1-3.

(a) The μ and σ values of the proposed features are highest for laughter followed by speech-laugh and neutral speech, for a given case. The higher values of the features in laughter are because of the more airflow through the vocal tract, making the vocal folds vibrate faster and stronger [7]. The presence of the continuous articulatory configuration for speaking in speech-laugh cuts the airflow, caused by co-occurring laughter, resulting in lower feature values compared to laughter, but still has higher feature values compared to neutral speech [10, 11].

(b) The variations observed in feature values between S_C and S_V follow a similar trend when compared to the corresponding variations in feature values observed between N_C and N_V. For instance, the μ of F_0, and the σ of F_0, β and γ are higher for S_C when compared with that of S_V, but the mean of β and γ are higher for S_V than S_C. Similarly, the μ of F_0, and the σ of F_0, β and γ are higher for N_C when compared with that of N_V, but the μ of β and γ are higher for N_V than N_C. It can also be observed that the μ and σ of all features are higher for S_C than that of N_C, and are also higher for S_V compared to that of N_V.

(c) It can be observed that the μ and σ of the feature values obtained for laughter, and the two cases of speech-laugh and neutral speech differ from each other. These variations in feature values are analyzed by finding the difference between the μ of feature values, and by computing the ratio of the σ of feature values for the cases $L\ vs\ S_C$, $L\ vs\ S_V$, $S_C\ vs\ N_C$ and $S_V\ vs\ N_V$, which are listed in Table 4.

In Table 4, Δ_F, Δ_β and Δ_γ refer to difference in μ of F_0, β and γ, respectively, obtained between two classes. For instance, Δ_F of $L\ vs\ S_C$ is $M_F(L)$ - $M_F(S_C)$, where M_F is the mean of F_0.

Table 2: Mean and standard deviation of feature values of speech-laugh and neutral speech for AMI dataset.

		S_C		S_V		N_C		N_V	
	Feature	μ	σ	μ	σ	μ	σ	μ	σ
Female	F_0	279.90	96.05	271.79	26.20	205.71	51.91	199.37	11.80
	β	83.83	51.90	94.50	35.08	75.94	41.06	77.15	19.37
	γ	31.37	23.94	37.04	13.85	21.25	12.88	25.47	4.75
Male	F_0	195.43	80.52	183.78	21.60	125.63	34.24	120.30	5.72
	β	61.60	37.21	66.93	21.39	45.74	21.78	50.25	8.52
	γ	15.22	12.41	16.08	4.69	8.89	5.21	9.26	1.61

Table 3: Mean and standard deviation of feature values of speech-laugh and neutral speech for SVL speech-laugh dataset.

		S_C		S_V		N_C		N_V	
	Feature	μ	σ	μ	σ	μ	σ	μ	σ
Female	F_0	298.03	97.31	286.00	29.37	218.14	55.28	208.31	10.72
	β	88.61	52.33	98.69	37.12	78.42	42.71	81.29	20.36
	γ	35.09	24.62	40.56	10.98	22.62	11.79	27.11	4.57
Male	F_0	204.95	82.18	198.63	20.98	138.36	38.45	136.93	6.05
	β	62.38	35.72	69.03	20.82	56.04	21.17	60.00	7.94
	γ	16.55	11.97	17.91	4.81	9.01	5.39	9.49	1.41

R_F, R_β and R_γ refer to the ratio of σ of the feature values of F_0, β and γ, respectively. For example, R_F of $L\ vs\ S_C$ is computed as the ratio of (σ_F of L) and (σ_F of S_C), where σ_F refers to σ of F_0.

It can be observed from Table 4 that even though, the difference in the μ of feature values (i.e., Δ_F, Δ_β and Δ_γ) obtained between laughter and speech-laugh are similar for both cases (i.e., vowels and complete utterance), the ratio of σ of the feature values (i.e., R_F, R_β and R_γ) obtained by considering vowels are higher compared to those obtained by considering complete utterance. Similar observation can be made from the values obtained between speech-laugh and neutral speech, which are given in Table 4.

These higher variations in ratio of σ of the feature values obtained by considering vowels can be exploited to discriminate speech-laugh from laughter and neutral speech.

4. Experimental Results and Discussion

Analysis of the features, as discussed in Section 3, shows that there are variations in difference of the μ of the feature values, and the ratio of σ of the feature values obtained by considering vowels and complete utterances of laughter, speech-laugh and neutral speech. These variations can be captured using Kullback-Leibler (KL) distance [17] which is computed as:

$$D = \frac{1}{2}(tr(\Sigma_1^{-1}\Sigma_0)+(\mu_1-\mu_0)^T\Sigma_1^{-1}(\mu_1-\mu_0)-k-ln(\frac{det\Sigma_0}{det\Sigma_1})), \quad (1)$$

where D is the KL distance, k is the dimension of the distribution, Σ_0, Σ_1 are covariance matrices and μ_0, μ_1 are the corresponding means of the 2-dimensional (2-D) distributions of reference and test utterances, respectively.

To compute KL distance, feature distributions of 2-D are formed by considering two features at a time. For the 3 proposed features, 3 feature distributions of 2-D i.e., (F_0 & β), (F_0 & γ) and (β & γ) are formed. Hence, each utterance is represented by these three feature distributions.

Table 5 gives the average KL distances obtained for the corresponding feature distributions between different combinations of the three classes. The average KL distances given in Table 5 are obtained by considering 4 female and 6 male speakers,

Table 4: Difference between the mean of features, and the ratio of SD of features obtained for AMI and SVL datasets.

		AMI				SVL speech-laugh			
	Feature	$L\ vs\ S_C$	$L\ vs\ S_V$	$S_C\ vs\ N_C$	$S_V\ vs\ N_V$	$L\ vs\ S_C$	$L\ vs\ S_V$	$S_C\ vs\ N_C$	$S_V\ vs\ N_V$
Female	Δ_F	60.16	**68.27**	**74.19**	72.42	62.31	**74.34**	**83.89**	77.69
	R_F	1.40	**5.16**	1.87	**2.22**	1.52	**5.05**	1.76	**2.74**
	Δ_β	**45.89**	35.22	7.89	**17.35**	**32.83**	22.75	10.19	**16.76**
	R_β	1.57	**2.32**	1.26	**1.81**	1.63	**2.30**	1.22	**1.82**
	Δ_γ	**38.75**	33.08	10.12	**11.57**	**34.16**	28.69	12.47	**13.45**
	R_γ	2.44	**4.22**	1.85	**2.91**	2.37	**5.35**	2.08	**2.40**
Male	Δ_F	75.24	**86.86**	**69.80**	63.48	65.69	**72.01**	**66.59**	61.70
	R_F	1.63	**6.08**	2.35	**3.77**	1.59	**6.26**	2.13	**3.46**
	Δ_β	**12.68**	7.35	15.86	**16.68**	**11.90**	5.25	6.34	**9.03**
	R_β	1.56	**2.73**	1.70	**2.51**	1.63	**2.80**	1.68	**2.62**
	Δ_γ	**13.81**	12.95	6.33	**6.82**	**12.48**	11.12	7.54	**8.42**
	R_γ	2.02	**5.36**	2.38	**2.91**	2.10	**5.22**	2.22	**3.41**

Table 5: Average KL distance values obtained among different feature distributions.

	Female			Male		
	F_0 & β	F_0 & γ	β & γ	F_0 & β	F_0 & γ	β & γ
$L_1\ vs\ L_2$	6.87	7.93	13.62	6.22	7.78	14.07
$L\ vs\ S_C$	13.81	13.45	17.94	9.59	20.72	26.41
$L\ vs\ S_V$	23.94	23.20	35.47	14.10	29.24	46.28
$N_{C1}\ vs\ N_{C2}$	7.35	9.12	10.53	6.42	7.39	9.43
$N_C\ vs\ S_C$	13.61	18.54	19.38	11.68	12.84	16.80
$N_{V1}\ vs\ N_{V2}$	4.16	4.89	6.03	2.36	2.91	3.74
$N_V\ vs\ S_V$	25.10	24.06	30.07	20.99	21.48	29.73

together from both AMI and SVL speech-laugh datasets. In Table 5, $L_1\ vs\ L_2$ refers to the case of computing the KL distance of the feature distributions between two different laughter segments of the same speaker. Similarly, $N_{C1}\ vs\ N_{C2}$ and $N_{V1}\ vs\ N_{V2}$ are the distances obtained between two different neutral utterances of the same speaker considering completing utterance and vowels, respectively. Based on the KL distance measures given in Table 5, thresholds are laid on each feature distribution pair as given in Table 6. Using these thresholds, experiments are performed to investigate the use of vowels in discriminating speech-laugh from laughter and neutral speech.

Table 6: Thresholds laid on feature distributions for experiment (Exp.) 1 and Exp. 2.

	C			V		
	F_0 & β	F_0 & γ	β & γ	F_0 & β	F_0 & γ	β & γ
Exp. 1	8.5	10	16	11	14	19
Exp. 2	9	12	13	9.5	10	12

4.1. Experiment 1

It can be observed from Table 5 that the KL distances obtained for $L_1 vs L_2$ are lower compared to those obtained between laughter and other segments (i.e., S_C and S_V). The KL distances obtained for $L\ vs\ S_V$ are higher compared to those obtained for $L\ vs\ S_C$. This shows that vowels rather than complete utterance of speech-laugh, provide better discrimination between speech-laugh and laughter. The steps in the experiment (Exp.) used to verify the use of vowels in discriminating speech-laugh from laughter are:

1. A sample laughter segment is taken as reference. A test segment (either laughter or speech-laugh) of the same speaker is taken for comparison with the reference.
2. All the three proposed features are extracted for both reference and test segments. In case of speech-laugh, features are extracted by considering two cases, i.e., complete utterance and vowels.
3. The KL distance for three 2-D distributions between reference and test segments are calculated, i.e., three KL distance measures are obtained for each case.
4. The KL distance obtained for each distribution is compared with the thresholds given in Table 6, for both cases.
5. If two or more KL distance values exceed the thresholds, then the test segment is considered as speech-laugh otherwise it is considered as laughter.

Table 7: Confusion matrix obtained by considering laughter as reference for AMI and SVL speech-laugh datasets.

		C		V	
		L	S	L	S
AMI	L	76.54%	23.46%	81.67%	18.33%
	S	24.13%	75.87%	19.21%	80.79%
SVL	L	77.28%	22.72%	83.11%	16.89%
	S	22.46%	77.54%	18.94%	81.06%

The results of Exp. 1 obtained on both datasets are given in Table 7. It can be observed from Table 7 that better discrimination between laughter and speech-laugh is obtained by considering vowels rather than the complete utterance of speech-laugh.

4.2. Experiment 2

It can be observed from Table 5 that the KL distances obtained for $N_{V1}\ vs\ N_{V2}$ are lower compared to those obtained for $N_{C1}\ vs\ N_{C2}$. Also, the KL distances obtained for $N_V\ vs\ S_V$ are higher compared to those obtained for $N_C\ vs\ S_C$. Exp. 2 is performed to verify the use of vowels in discriminating speech-laugh from neutral speech. The steps in Exp. 2, are same as the steps followed for Exp. 1, except that the reference utterance considered for Exp. 2 is neutral speech of a speaker, but not laughter, and the discrimination is between speech-laugh and neutral speech. The thresholds given in Table 6 are used for Exp. 2.

Table 8 provide the results of Exp. 2 obtained for AMI and SVL speech-laugh datasets. It can be observed from Table 8 that the discrimination between neutral speech and speech-laugh is better in the case of considering vowels instead of considering the complete utterance, for both datasets.

Table 8: Confusion matrix obtained by considering neutral speech as reference for AMI and SVL speech-laugh datasets.

		C		V	
		N	S	N	S
AMI	N	79.59%	20.41%	85.79%	14.21%
	S	25.03%	74.97%	15.39%	84.61%
SVL	N	78.45%	21.55%	84.98%	15.02%
	S	27.86%	72.14%	16.84%	83.16%

5. Summary and Conclusions

In this paper, the use of excitation source based features extracted in vowel regions were analyzed for discriminating speech-laugh from laughter and neutral speech. The features based on fundamental frequency and strength of excitation at epochs were extracted using modified zero frequency filtering method. Analysis of these features shows that the variations between speech-laugh, and laughter and neutral speech are higher by considering vowels rather than the complete utterance of speech-laugh and neutral speech. This observation was exploited to achieve better discrimination of speech-laugh from laughter and neutral speech. Further work need to be done to investigate the use of vowels in detecting speech-laugh segments in conversational speech.

6. References

[1] Caroline Menezes and Yosuke Igarashi, “The speech laugh spectrum,” in *Proc. Speech Production*, Brazil, pp. 157-164, Dec. 2006.

[2] E. E. Nwokah, Hui-Chin Hsu, P. Davies and A. Fogel, “The integration of laughter and speech in vocal communication: a dynamic system perspective,” *Journal of Speech, Language and Hearing Research*, vol. 42, no. 4, pp. 880-894, Aug. 1999.

[3] J. Bachorowski, M. J. Smoski and M. J. Owren, “The acoustic features of human laughter,” *Journal of the Acoustical Society of America*, vol. 110, no. 3, pp. 1582-1597, Jun. 2001.

[4] Jieun Oh, Eunjoon Cho and Malcolm Slaney, “Characteristic contours of syllabic-level units in laughter,” in *Proc. Interspeech*, Lyon, France, pp. 158-162, Aug. 2013.

[5] khiet P. Truong and Jürgen Trouvain, “Investigating prosodic relations between initiating and responding laughs,” in *Proc. Interspeech*, Singapore, pp. 1811-1815, Sep. 2014.

[6] Teun F. Krikke and khiet P. Truong, “Detection of nonverbal vocalizations using Gaussian Mixture Models: looking for fillers and laughter in conversational speech,” in *Proc. Interspeech*, Lyon, France, pp. 163-167, Aug. 2013.

[7] K. Sudheer, M. Sri Harish Reddy, K. Sri Rama Murty and B. Yegnanarayana, “Analysis of laugh signals for detecting in continuous speech,” in *Proc. Interspeech*, Brighton, UK, pp. 1591-1594, Sep. 2009.

[8] Carlos Ishi, Hiroaki Hatano and Norihiro Hagita, “Analysis of laughter events in real science classes by using multiple environment sensor data,” in *Proc. Interspeech*, Singapore, pp. 1043-1047, Sep. 2014.

[9] L. S. Kennedy, and D. P. w. Ellis, “Laughter detection in meetings,” in *Proc. NIST ICASSP 2004 Meeting Recognition Workshop*, Montreal, Canada, pp. 118-121, May 2004.

[10] Jürgen Trouvain, “Phonetic aspects of speech laughs,” in *Proc. ORAGE*, Paris, France, pp. 634-639, Jun. 2001.

[11] Sri Harsha Dumpala, Karthik Venkat Sridaran, Suryakanth V. Gangashetty, and B. Yegnanarayana, “Analysis of laughter and speech-laugh signals using excitation source information,” in *Proc. ICASSP*, Florence, Italy, pp. 975-979, May 2014.

[12] V. K. Mittal and B. Yegnanarayana, “Study of changes in glottal vibration characteristics during laughter,” in *Proc. Interspeech*, Singapore, pp. 1777-1781, Sep. 2014.

[13] V. K. Mittal and B. Yegnanarayana, “Analysis of production characteristics of laughter,” in *Computer Speech and Language*, vol 30, no 1, pp. 99-115, Mar. 2015.

[14] Jieun Oh and Ge Wang, “Laughter modulation: from speech to speech-laugh,” in *Proc. Interspeech*, Lyon, France, pp. 754-755, Aug. 2013.

[15] I. McCowan, J. Carletta, W. Kraaij, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, W. Post, D. Reidsma, and P. Wellner, “The AMI meeting corpus,” in *Proc. Conference on Methods and Techniques in Behavioral Research*, vol. 88, pp. 137-140, Sep. 2005.

[16] Sri Rama Murty K., B. Yegnanarayana and Anand Joseph Xavier M., “Characterization of glottal activity from speech signals,” *IEEE Signal Processing Letters*, vol. 16, no. 6, pp. 469-472, Jun. 2009.

[17] J. R. Hershey and P. A. Oslen, “Approximating the Kullback-Leibler divergence between Gaussian mixture models,” in *Proc. ICASSP*, Honolulu, Hawaii, USA, pp. 317-320, Apr. 2007.