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

Laughter Valence Prediction in Motivational Interviewing based on Lexical and Acoustic Cues

Rahul Gupta[o], Nishant Nath[*], Taruna Agrawal[o],
Panayiotis Georgiou[*], David Atkins[+], Shrikanth Narayanan[o]

[o]Signal Analysis and Interpretation Lab (SAIL),
[*]Signal Processing for Communication Understanding and Behavior Analysis (SCUBA),
University of Southern California, Los Angeles, USA
[+]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, USA

Abstract

Motivational Interviewing (MI) is a goal oriented psychotherapy counseling that aims to instill positive change in a client through discussion. Since the discourse is in the form of semi-structured natural conversation, it often involves a variety of non-verbal social and affective behaviors such as laughter. Laughter carries information related to affect, mood and personality and can offer a window into the mental state of a person. In this work, we conduct an analytical study on predicting the valence of laughters (positive, neutral or negative) based on lexical and acoustic cues, within the context of MI. We hypothesize that the valence of laughter can be predicted using a window of past and future context around the laughter and, design models to incorporate context, from both text and audio. Through these experiments we validate the relation of the two modalities to perceived laughter valence. Based on the outputs of the prediction experiment, we perform a follow up analysis of the results including: (i) identification of the optimal past and future context in the audio and lexical channels, (ii) investigation of the differences in the prediction patterns for the counselor and the client and, (iii) analysis of feature patterns across the two modalities.

Index Terms: Laughter valence, Context analysis, Multi-modal classification and fusion, Behavioral signal processing.

1. Introduction

Motivational interviewing (MI) [1] is a psychotherapeutic intervention for substance abuse involving dialog between a counselor and a client (the patient). The counselor attempts to motivate the client towards positive behavior, i.e. against addictive behavior, in a semi-structured conversational format. The conversation includes non-verbal expressions including laughter, sighs, facial expressions and body gestures. In particular, laughter has been widely studied in human conversation [2,3] and has been a subject of our previous investigations [4, 5] within the MI protocol. Arguably, laughter is often associated with affective expression, the understanding of which is of importance in psychotherapy [6]. In this work, we investigate the affective expressions associated with laughters in MI sessions, specifically their valence. We initially model the information predictive of laughter valence in lexical and acoustic channels by developing a classification scheme for the same. This is followed by model analysis and we make investigations on the optimal context length in the lexical and acoustic channels for valence prediction, model performance across the two speaker groups (i.e., counselor and client) and, the most important lexical and acoustic features related to laughter valence. Our overarching goal in this work is to enhance the understanding of laughter phenomenon within the MI protocol, thus aiding a more effective intervention.

Past work has investigated laughters in relation to emotions [7, 8], nonlinguistic communication [9] and pathology [10]. Laughter has also been a subject of investigation in several psychotherapy studies [6] including in MI studies [11]. Since laughter is a multi-modal event, researchers have further looked into multi-modal modeling schemes for laughters. For instance, Melder et al. [12] developed a multi-modal mirror that senses user states and elicits laughters. Multi-modal studies of laughter have led to precise detection [13], developing interactive systems [14] and, supporting emotion analysis [15]. Also within the domain of MI (the subject of this paper), researchers have investigated the role of laughters [4, 16]. Despite this, a comprehensive multi-modal analysis of laughters is still lacking. We approach this issue in this paper by performing an analysis of laughters using language and acoustic information.

We work with a set of MI protocol based clinical trials and annotate laughters in terms of conveying a positive, neutral or negative valence. We then develop two systems based on lexical and acoustic cues, respectively, for the prediction of laughter valence, followed by a system fusion. The goal of these experiments is to demonstrate that both lexical and acoustic cues are associated with laughter valence. Based on the outputs of the valence prediction system, we perform a set of three analyses on: (i) computing the optimal past and future context in the two modalities for valence prediction, (ii) evaluating model performance conditioned on the speaker group (client or counselor) and, (iii) analyzing top features in the lexical and acoustic streams contributing to laughter valence prediction. The analysis reveals the signature of laughter valence is encoded over a longer past context than the future, and that prosodic features are more discriminative of laughter valence than spectral features. In the next section, we describe our dataset followed by the description of the experimental methodology in Section 3.

2. Dataset

For this study, we use a set of 92 sessions from five MI clinical trials namely: HMCBI, ESPSB, ESP21, ARC and iCHAMP sessions [4]. All these sessions contain conversations between a counselor and a patient discussing substance (such as alcohol, drugs and tobacco) abuse. Each of these sessions are segmented at an utterance level by a specialist trained using the Motivational Interviewing Skill Code (MISC) manual [17]. The MISC manual has a five point definition for an utterance including criteria such as an utterance should be a complete thought and should have speaker continuity. Excerpts of this dataset can be found in our previous works [4,5].

In the set of 92 sessions, we observe a count of 1291 laughters with 597 of these laughters belonging to the counselor. We

http://dx.doi.org/10.21437/Interspeech.2016-184

Table 1: Example of utterance containing laughter from each of the classes

Class	Example utterance	Comments
Positive	Client: I probably won't drink with my family Counselor: Me neither **[laughs]**	Shared enjoyment
Neutral	Counselor: So people act up ? Client: Yeah, that is stupid **[laughs]**	Conversational
Negative	Client: I do not think I can do it **[laughs]**	Self-pity

Table 2: Statistics for laughter annotations for both the speakers

Speaker	Count			
	Positive	Neutral	Negative	Total
Counselor	136	453	8	597
Client	124	483	87	694
Combined	260	936	95	1291

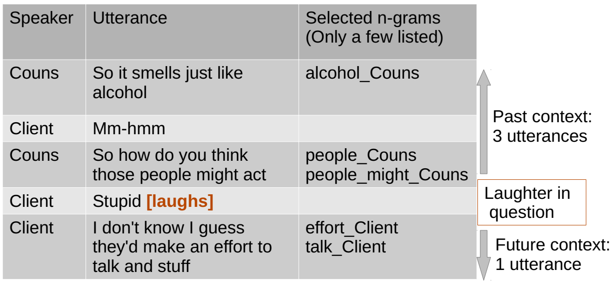
annotate each of these laughters as carrying a positive, negative or neutral valence. The definition of these labels is inspired from the existing literature. Provine [3] and Glenn [2] in their books discuss various categories of laughters including contagious, inappropriate, abnormal and equivocal laughters. Several researchers in machine learning and signal processing have also focused on laughter classification based on the emotion content. For instance, Szameitat et al. [18] classify laughters into four categories (joy, tickling, taunting, schadenfreude) based on the underlying emotions. Similarly, Miranda et al. [19] classify laughters into one of five categories of emotions specific to the Filipino culture. In this work, the classes of laughter are defined based on the perceived emotion carried by the laughter. Positive laughters include laughters used to express happiness, excitement, (shared) enjoyment and/or pleasure. Neutral laughters are defined to be conversational laughters and are often used as a placeholder during conversations. Negative laughters are accompanied by utterances that reflect embarrassment, self pity, discomfort and/or sarcasm. Examples of each category of laughter along within the MI framework are shown in Table 1. These annotations were carried out by the second author of this paper in discussion with the first author. Statistics of the laughter labels are shown in Table 2. We would like to point out that the most significant difference between counselor and client laughters is in the occurrence of negative laughters, with far less proportion of negative laughters for the counselor. It is expected that counselor instills a positive motivation in the client so the negative valence counselor laughters should be minimized.

3. Experiments

We divide our experiments in two parts. We initially perform a classification experiment to identify the laughter valence based on lexical and acoustic cues. This is followed by the analysis of model parameters, specifically context length of lexical and acoustic cues, and performances for the psychologist and the client side of the interaction.

3.1. Prediction of laughter valence

In this experiment, we design models to predict the annotated laughter valence based on lexical and acoustic cues. The goal is to validate if these cues carry information regarding the perceived valence of laughter instead of accuracy focused automation of affect prediction. The reasons for choosing the former as an objective is that perception of behavioral attributes (including laughter valence) is diverse and subjective across the population and conditioned on the context. Therefore accuracy driven models warrant the use of methods accounting for this important attribute of human behavioral data; examples include mixture of experts models [20], multiple annotator models [21] and models accounting for human factors such as reliability [22]. Our model is instead focused on validating if there are any patterns in the lexical and acoustic channels with regards to the perceived laughter valence.



Speaker	Utterance	Selected n-grams (Only a few listed)
Couns	So it smells just like alcohol	alcohol_Couns
Client	Mm-hmm	
Couns	So how do you think those people might act	people_Couns people_might_Couns
Client	Stupid [laughs]	
Client	I don't know I guess they'd make an effort to talk and stuff	effort_Client talk_Client

Figure 1: Example of extracting n-grams based on counselor and client utterances for training the MaxEnt classifier. In this specific example the past/future context lengths are 3/1 utterances. Note that the speaker role is appended to each n-gram.

We design two separate models for each of these channels followed by a weighted fusion scheme. The models are trained on combined data from counselor and client laughters due to two reasons: (i) firstly, to develop a prediction model universal to both the speaker groups and, (ii) secondly, a model trained on combined data has more data samples to train on. For each of the prediction experiments, we perform a 10 fold cross-validation with 80% of the data used as a training set, 10% as the development set and the remaining 10% as the test set. We chose the Unweighted Average Recall (UAR) as the evaluation metric for the classification system due to unbalanced distribution of instances among the positive, negative and neutral classes. UAR was the metric of choice for several other experiments with imbalance in the data [23, 24]. We describe the experiments for valence prediction below.

3.1.1. Prediction of laughter valence based on lexical cues

In this experiment, we predict laughter valence based on the utterances around the laughter. Given a window of utterances from the past as well as the future, we compute a set of unigrams and bigrams for each of the utterances. The n-grams are further appended with the speaker role tag to carry information regarding the source of the n-gram. The n-grams from the training set are then used to train a Maximum Entropy (MaxEnt) classifier with target labels as the laughter valence. Due to a large feature dimensionality associated with the n-grams, we prune n-grams based on a minimum count of occurrence, tuned on the development set. A schematic of an utterance window along with extracted n-grams is shown in Figure 1. The length of the utterance window in the future and in the past is also tuned on the development set for each iteration. This window length is agnostic to the count of utterances from individual speakers within the window. Hence the window can contain any number of utterances from the two speakers as long as the number of utterances sum to the window length. We did not chose a window length for each speaker individually as it leads to a longer context from the past/future, in the case of unbalanced conversations when one speaker speaks more than the other. Algorithm 1 presents a summary of the training algorithm for classification based on lexical cues.

3.1.2. Prediction of laughter valence based on acoustic cues

Following the classification setup in the previous section based on lexical cues, we also perform laughter valence prediction

Algorithm 1 Summary of training procedure for classification based on lexical cues.

```
Select a number of future and past utterances around the utterance containing laughter (tuned on the development set).
Extract unigrams and bigrams from utterances with speaker role appended to the n-grams.
Select n-grams that have a count higher than a threshold (threshold also tuned on the development set).
Train a MaxEnt model on the selected n-grams for predicting positive, neutral and negative laughter valence classes.
```

Table 3: Acoustic-prosodic signals and statistical functionals computed over them for the classification experiment using acoustic cues

Acoustic-prosodic signals	Mel-Frequency Cepstral Coefficients (MFCC), F0, harmonic to noise ratio, intensity, zero-crossing rate, (+Δ + $\Delta\Delta$ for all signals)
Statistical functionals	Mean, median, Inter-Quantile Ratio (IQR), standard deviation

based on acoustic cues. The setup of this experiment is inspired from the work by Chaspari et al. [25] for classification of social laughters. Given a laughter location from a speaker, we first extract a few statistical functionals on acoustic prosodic signals from a segment containing that laughter. Apart from the laughter, the segment also contains a past/future context, length of which is again tuned on the development (with steps of 30 milliseconds). An illustration of a segment from the client is shown in Figure 2. The acoustic prosodic signals and the statistical functionals used in the experiment are listed in Table 3 and are extracted using the OpenSMILE software [26]. The acoustic-prosodic signals are z-normalized per speaker and statistical functionals are computed only on frames with voicing probability (also computed using OpenSMILE) greater than 0.5. We limit to only a few statistical functionals to limit the feature dimensionality during classification. This is desirable as our dataset contains a limited number of samples. Classification is performed using a linear Support Vector Machine (SVM) classifier with the complexity parameter C [27] tuned on the development set.

3.1.3. Fusion of lexical and acoustic cue based systems

In order to fuse the valence prediction results obtained from the lexical and acoustic cues, we perform a weighted fusion of probabilities from the two systems. Let the positive valence class probability for a laughter, as output by the MaxEnt classifier trained on lexical cues, be p_l^+. Similarly the probability from the SVM classifier (computed by fitting logistic model to distances from class hyperplanes [28]) based on the acoustic cues for the positive class is represented as p_a^+. The final fusion score p_f^+ for the positive class is computed as shown in (1). α is the weighting parameter tuned on the development set.

$$p_f^+ = \alpha p_l^+ + (1-\alpha)p_a^+ \qquad (1)$$

Similarly, we compute the fusion scores for the negative and neutral classes. The final class assignment is computed by scaling the system probabilities/ fusion scores by class frequencies as discussed in the next section. We also present the results and discuss the findings in the following section.

3.1.4. Results and discussion

Given that the models are trained on unbalanced data, we inversely scale the system probabilities/ fusion scores for each class with the instance count for that class in the training set. The final class assignment is the one with the highest scaled probability/fusion score. This provides a more balanced recall per class in the computation of UAR. The UAR for lexical, acoustic and fusion systems is shown in Table 4.

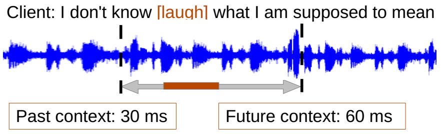


Figure 2: Example of extracting acoustic cues from speech. In this example, we trim the speech starting at 30 milliseconds before laughter begins and 60 milliseconds after laughter ends. This is followed by extraction of prosodic signals and computation statistical functionals on the speech segment.

Table 4: Unweighted Average Recall (UAR) for the lexical and acoustic classification system and their fusion

System	UAR	Class recalls in %		
	(in %)	Positive	Neutral	Negative
Chance	33.3	-	-	-
Lexical	45.7	38.8	39.4	58.9
Acoustic	38.1	39.2	41.2	33.7
Fusion	**46.2**	39.2	39.3	60.0

From the results, we observe that the UAR for classification based on lexical cues is significantly better than chance with p-value $< 5\%$ (binomial proportions test). However the UAR for classification based on acoustic cues is slightly weaker and significantly better than chance with p-value $< 10\%$ (binomial proportions test). Both these results show that there exists information in both lexical and acoustic channels for identifying the valence content of the laughter. However, the acoustic channel is weaker in prediction and its fusion with the lexical cues system outputs marginally improves the overall UAR to 46.2% from 45.7% (this improvement is not significant). This is due to the fact that there were only a handful of instances where the system based on acoustic cues made the right prediction and system based on lexical cues did not. Nevertheless, the combined system performs the best and encourages further investigation into the joint performance of acoustic and lexical cues. We discuss a few implications of the system in the next section.

3.2. Analysis of model parameters and outputs

In this section, we perform three sets of analyses on the classification model and parameters presented in the last section: (i) optimal context length for the lexical and acoustic classification systems, (ii) individual model performance for counselor and client laughters and, (iii) a feature analysis for the lexical and acoustic cue based systems. We discuss each of these experiments below.

3.2.1. Optimal context length for the classification systems

In this section, we investigate the optimal context length for the lexical and acoustic systems, determined empirically, and comment on the patterns. We rerun the classification experiment for both the modalities; however, instead of tuning the past/future context lengths based on the development set at each cross-validation iteration, the context lengths are kept constant. This is performed to determine which context lengths universally provide the best classification accuracy on the entire data. The UAR for each context length combination (future and past) in the lexical and acoustic systems is shown as a matrix image in Figure 3 (top) and Figure 3 (bottom), respectively.

From the figures we observe that a longer past context

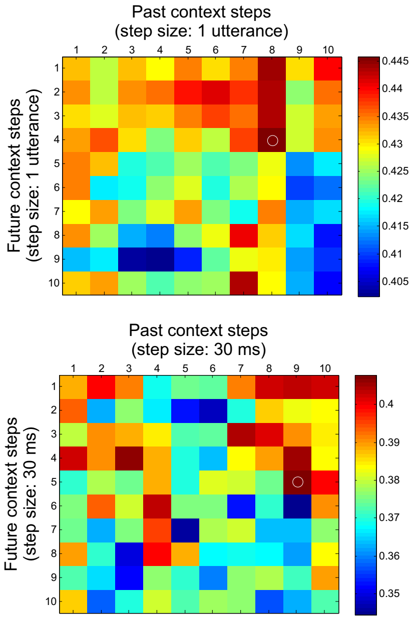


Figure 3: UARs obtained from the lexical (top) and acoustic (bottom) cue based classification systems with the past/future context size fixed during cross-validation. Colorbar on right shows the UAR values. The cell with a white circle is the best performance in the two matrices.

Table 5: Unweighted Average Recall (UAR) after fusion for each speaker group. We also mention the class counts in brackets along with the performance of fusion system in the last row.

Speaker	UAR	Class recalls in % (counts)		
	(in %)	Positive	Neutral	Negative
Counselor	36.7	58.8(136)	51.2(453)	0.0(8)
Client	37.1	17.7(124)	28.1(483)	65.1(87)
Combined	46.2	39.2 (260)	39.3 (936)	60.0 (95)

(compared to future context) provides the best UAR for both the modalities. This implies that the laughter valence is reflected over a longer context in the past (than in the future) in case of both the modalities. The matrix for classification based on lexical cues (Figure 3 (top)) appears to be more structured with high values around the cell with the highest UAR. This indicates a smooth decay of information regarding laughter valence around that particular context length. Although, a few cells around the optimal cell in the matrix for classification based on acoustic cues (Figure 3 (bottom)) also carry high values, the pattern is more noisy. For instance, fourth row - third column and sixth row - fourth column also carry a high values of UAR for the acoustic system. This suggests that the acoustic cue based classification system is more noisy in determination of optimal context. This observation is consistent with the results in Table 4 with a lower performance using the acoustic cues.

3.2.2. *Model performance for counselor and client laughters*

As previously mentioned, we trained a model universal to the two groups of speakers. In this section, we investigate how well the model generalizes to the groups individually. Table 5 presents the UARs for counselor and client laughters separately, as computed after the fusion of lexical and acoustic systems.

From the results in Table 5, we observe that the UAR performances for both the speaker groups are close, however the per-class accuracies are significantly different across the two groups (binomial proportions test, p-value $< 5\%$). It is interesting to note that the class patterns captured by the model are conditioned on the speaker group. For example, a high class recall for the positive class within the counselor group suggests that it is easy to discriminate a positive counselor laughter based on acoustic and lexical cues. However, the same is not true for the client group with lower recall for the positive class. Next, we list top few cues associated with the classification system.

Table 6: Top lexical and acoustic cues for classification. The class within brackets for lexical cues shows the class favored by the n-gram.

Top n-grams	Top acoustic-prosodic statistical functionals
risks_couns (neutral)	IQR: F0
good_so_couns (neutral)	Median: F0
is_expensive_client (neutral)	Median: intensity
outgoing_couns (positive)	Median: ΔHNR
had_not_client (negative)	Median: HNR

3.2.3. *Feature analysis for lexical and acoustic systems*

In this section we list the top five n-grams and acoustic-prosodic statistical functionals associated with laughter valence classes. The top n-grams are the ones that have the highest output probability (inversely scaled by count of class instances) favoring any one of the three valence classes, as determined by the MaxEnt classifier. The top acoustic-prosodic statistical functionals are computed based on their mutual information with the valence classes. Table 6 shows the top cues for the lexical and acoustic systems.

We observe several interesting feature patterns in Table 6. The n-grams can be weakly associated with the class they correspond to. For instance, the word “outgoing” uttered by counselor can be associated with positive emotions and hence associates with the positive class. Similarly, the bigram “had_not” uttered by client could be associated with retrospection or regret, hence associating with the negative class. Another interesting observation is that the top acoustic features are all prosodic features with F0 being part of top two features. Although we also extract MFCCs (which reflect spectral properties of speech), they are not present in the top features. This indicates that prosody carries substantial information in perception of laughter valence.

4. Conclusion

Motivational Interviewing (MI) is a goal oriented psychotherapy with semi-structured conversations between a counselor and a client, which often includes laughter as a mode of expression. In this work, we analyzed the emotion content of laughters by developing a classification scheme based on lexical and acoustic cues. We showed that these two channels contain information regarding the laughter valence and performed follow up analysis. We investigated the role of context and observed that a longer past context carries information regarding laughter valence than the future context. We also commented on classification accuracies per speaker group and identified top features important for classification.

In the future, we aim at performing further analysis with finer annotations on laughters within MI, which incorporate the dimensions of arousal and dominance in addition to valence. Since the perception of emotions is observer dependent, we also aim on training multiple annotator models such as the one proposed by Raykar et al. [21]. Finally, the study could also be extended to other non-verbal cues such as sighs, body gestures and facial expressions.

5. References

[1] S. Rollnick and W. Miller, "What is motivational interviewing?" *Behavioural and cognitive psychotherapy*, vol. 23, no. 04, pp. 325–334, 1995.

[2] P. Glenn, *Laughter in interaction*. Cambridge University Press, 2003, vol. 18.

[3] R. R. Provine, *Laughter: A scientific investigation*. Penguin, 2001.

[4] R. Gupta, T. Chaspari, P. G. Georgiou, D. C. Atkins, and S. S. Narayanan, "Analysis and modeling of the role of laughter in motivational interviewing based psychotherapy conversations," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[5] R. Gupta, P. G. Georgiou, D. C. Atkins, and S. S. Narayanan, "Predicting client's inclination towards target behavior change in motivational interviewing and investigating the role of laughter." in *INTERSPEECH*, 2014, pp. 208–212.

[6] A. R. Mahrer and P. A. Gervaize, "An integrative review of strong laughter in psychotherapy: What it is and how it works." *Psychotherapy: Theory, Research, Practice, Training*, vol. 21, no. 4, p. 510, 1984.

[7] M. J. Owren and J.-A. Bachorowski, "The evolution of emotional experience: A" selfish-gene" account of smiling and laughter in early hominids and humans." 2001.

[8] N. A. Kuiper and R. A. Martin, "Laughter and stress in daily life: Relation to positive and negative affect," *Motivation and Emotion*, vol. 22, no. 2, pp. 133–153, 1998.

[9] M. J. Owren and J.-A. Bachorowski, "Reconsidering the evolution of nonlinguistic communication: The case of laughter," *Journal of Nonverbal Behavior*, vol. 27, no. 3, pp. 183–200, 2003.

[10] D. W. Black, "Pathological laughter: a review of the literature." *The Journal of nervous and mental disease*, vol. 170, no. 2, pp. 67–71, 1982.

[11] E. McNamara, "Motivational interviewing and cognitive intervention," *Working with emotions: responding to the challenge of difficult pupil behaviour in schools*, pp. 77–98, 2001.

[12] W. A. Melder, K. P. Truong, M. D. Uyl, D. A. Van Leeuwen, M. A. Neerincx, L. R. Loos, and B. Plum, "Affective multimodal mirror: sensing and eliciting laughter," in *Proceedings of the international workshop on Human-centered multimedia*. ACM, 2007, pp. 31–40.

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

[14] J. Urbain, R. Niewiadomski, M. Mancini, H. Griffin, H. Çakmak, L. Ach, and G. Volpe, "Multimodal analysis of laughter for an interactive system," in *Intelligent Technologies for Interactive Entertainment*. Springer, 2013, pp. 183–192.

[15] M. T. Suarez, J. Cu, and M. Sta, "Building a multimodal laughter database for emotion recognition." in *LREC*, 2012, pp. 2347–2350.

[16] H. A. Westra and A. Aviram, "Core skills in motivational interviewing." *Psychotherapy*, vol. 50, no. 3, p. 273, 2013.

[17] W. R. Miller, T. B. Moyers, D. Ernst, and P. Amrhein, "Manual for the motivational interviewing skill code (MISC)," *Unpublished manuscript. Albuquerque: Center on Alcoholism, Substance Abuse and Addictions, University of New Mexico*, 2003.

[18] D. P. Szameitat, K. Alter, A. J. Szameitat, C. J. Darwin, D. Wildgruber, S. Dietrich, and A. Sterr, "Differentiation of emotions in laughter at the behavioral level." *Emotion*, vol. 9, no. 3, p. 397, 2009.

[19] M. Miranda, J. A. Alonzo, J. Campita, S. Lucila, and M. Suarez, "Discovering emotions in filipino laughter using audio features," in *3rd International Conference on Human-Centric Computing (HumanCom)*. IEEE, 2010, pp. 1–6.

[20] S. Gutta, J. R. Huang, P. Jonathon, and H. Wechsler, "Mixture of experts for classification of gender, ethnic origin, and pose of human faces," *Neural Networks, IEEE Transactions on*, vol. 11, no. 4, pp. 948–960, 2000.

[21] V. C. Raykar, S. Yu, L. H. Zhao, G. H. Valadez, C. Florin, L. Bogoni, and L. Moy, "Learning from crowds," *The Journal of Machine Learning Research*, vol. 11, pp. 1297–1322, 2010.

[22] N. Kumar and S. Narayanan, "A discriminative reliability-aware classification model with applications to intelligibility classification in pathological speech," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[23] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, F. Burkhardt, R. Van Son, F. Weninger, F. Eyben, T. Bocklet *et al.*, "The INTERSPEECH 2012 speaker trait challenge." in *INTERSPEECH*, vol. 2012, 2012, pp. 254–257.

[24] R. Gupta, C.-C. Lee, and S. Narayanan, "Classification of emotional content of sighs in dyadic human interactions," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 2265–2268.

[25] T. Chaspari, E. M. Provost, A. Katsamanis, and S. Narayanan, "An acoustic analysis of shared enjoyment in ECA interactions of children with autism," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 4485–4488.

[26] F. Eyben, M. Wöllmer, and B. Schuller, "OpenSMILE: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 1459–1462.

[27] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[28] T. Hastie, R. Tibshirani *et al.*, "Classification by pairwise coupling," *The annals of statistics*, vol. 26, no. 2, pp. 451–471, 1998.