DIMENSIONAL ANALYSIS OF LAUGHTER IN FEMALE CONVERSATIONAL SPEECH

Mary Pietrowicz[1,2], Carla Agurto[1], Jonah Casebeer[2], Mark Hasegawa-Johnson[3], Karrie Karahalios[2], and Guillermo Cecchi[1]

IBM[1], and University of Illinois, Departments of Computer Science[2] and Electrical and Computer Engineering[3]

ABSTRACT

How do people hear laughter in expressive, unprompted speech? What is the range of expressivity and function of laughter in this speech, and how can laughter inform the recognition of higher-level expressive dimensions in a corpus? This paper presents a scalable method for collecting natural human description of laughter, transforming the description to a vector of quantifiable laughter dimensions, and deriving baseline classifiers for the different dimensions of expressive laughter. Then, it explores the impact of leveraging nuances of laughter in the recognition of higher-level, general expressive dimensions, discovered in the same way, such as genuine happiness, sarcasm, nervous reflection, and more. The performance of the low-level laughter classifiers is presented, along with the performance of the high-level laughter-aware and laughter-unaware classifiers.

Index Terms— laughter, perception, vocal expression, latent semantic analysis, dimensional analysis

1. INTRODUCTION

Although laughter is associated with humor, human laughter signals a range of other emotions and intentions, and serves many different social functions, such as signaling of empathy, attention, agreement, approval, sarcasm, social dominance, and connection. It also provides a mechanism for emotional regulation, a type of "relief valve," which is the reason we laugh in moments of negative affect, such as nervousness, stress, and sadness. These laughter types sound different (imagine the belly laugh vs. a sarcastic snicker vs. a breathy 'heh'), and are processed differently in the brain [3,25,26]. Most of the current research on laughter, however, does not explore this range and function of laughter, and considers it a single type of utterance. These studies detect the presence of laughter in speech via acoustic analysis, feature selection, and machine modeling, and are often motivated by the need to improve speech recognition [6,9,10,11,13,18]. Similar work analyzes laughter linguistically, in the context of prosody [1,19]. This approach is more closely related to expressive purpose, but the result does not distinguish one kind of laughter from another. Explorations which make distinctions among laughter types typically limit it to speech, laughter, joint laughter, and speech-laugh [4,17,20,27]. A typical approach to studying the meaning of laughter focuses on a specific laughter type, for example, the relationships among speech-laugh, parentese, and specific emotions [2]. Other studies explore meaning via the social dimensions of laughter, for example, distinguishing solo laughter from joint laughter, distinguishing initiating laughter from responding laughter, and showing the impact of interaction on the laughter rate [28]. A few of these explorations show how laughter can be used to detect specific higher-level social behaviors, such as empathy, acceptance, collaboration [7], and the attitude toward a proposed behavior change [8]. A final category of laughter research produces interactive simulations involving laughter, and studies how humans respond to robots which detect and produce laughter in different scenarios [2,16,29].

Our work extends the range of laughter types previously explored by 1) leveraging descriptions of what people hear in expressive laughter, 2) applying dimensional discovery of perceived laughter types across a corpus of oral history interviews, and 3) producing models which can evaluate laughter samples with respect to the newly-discovered laughter types, resulting in a descriptive laughter vector. Next, we show how the resulting laughter vector can be used to classify the expressive modality of the entire containing utterance. The following research questions guided our work:

RQ1: What perception-grounded dimensions of laughter can be found in conversational speech?
RQ2: How can these discovered dimensions of laughter be modeled acoustically?
RQ3: How can the resulting laughter models be used to recognize other dimensions of vocal expression?

2. ORAL HISTORY CORPUS

The library of congress Veterans' Oral History Project [15] is a freely-accessible collection of semi-structured conversational speech. The interview context provides a common structure across all samples, and the question content is strikingly similar across the corpus. Most interviewers, for example, asked subjects to state their

Authorized licensed use limited to: University of Illinois. Downloaded on April 27,2022 at 15:50:03 UTC from IEEE Xplore. Restrictions apply.

names, provide demographic information, explain their reasons for joining the military, describe their military training, and tell one or more personal stories. The similarities in structure and question provided a natural baseline for comparison across interviews, and the personal experiences provided a diverse range of laughter to explore. Speakers laughed when recalling or responding to a range of emotional triggers, including absurdity, surprise, humor, tension, nervousness, sorrow, fear, annoyance, and despair, to name a few. They also laughed to express sarcasm and for conversational connection and flow. We focus on female speakers here (both interviewees and interviewers), leverage prior work in the discovery and modeling of general expressive dimensions in this corpus [24], and relate general dimensions of vocal expressivity to the specific dimensions of laughter. The speaker selections from the prior study [24] were used to study laughter here.

3. ANALYSIS OF PERCEPTION

This section describes the analysis of the perception of laugher [22], reviews the analysis of the perception of general expressivity from a prior study [24], and shows the relationship between the two.

3.1. Perception of Laughter

To begin to analyze laughter and address RQ1, we extracted 120 laughter episodes from ten representative female (for acoustic similarity) talkers, and tagged each laughter event with one of four laughter types, which included single-person laughter events (laughing alone: 60%, and simultaneous laugh-speech: 14%) and interactive laughter (joint laughter: 15%, and joint laughter-speech: 11%). Then, we asked ten Mechanical Turk workers per laughter sample to provide three or more keywords describing the expressivity in the laughter, for a total of 1200 Turk tasks and 4000+ descriptive keywords. The range and distribution of keyword descriptors given for laughter is similar to those provided when listeners are asked to describe general speech expressivity [24]. Over half of the laughter keywords describe emotion in a nuanced way, far beyond the range of any theory of basic emotion [21]. About 40% describe the combined prosody of laughter (a narrow collection of speed, duration, loudness, and articulation words) and voice quality. The voice qualities for laughter include many of the same descriptors given for general vocal expressivity, such as breathiness and resonance; in addition they include a new vocabulary of laughter-specific qualities, such as "chuckle," "giggle," "chortle," and "snort." The remainder of the keywords are attributed personal qualities, such as sincerity. To discover expressive dimensions of laughter from listener perception, and to explore relationships among perceived keyword qualities, we performed latent semantic analysis (LSA) [14] across the keywords and laughter clips to discover the expressive dimensions of laughter. The LSA technique begins with a matrix of descriptor counts per audio clip, and applies singular value decomposition (SVD) to this matrix, ultimately resulting in weighted associations between the descriptors/audio clips and the discovered LSA dimensions (which are described by the weighted association of human-generated keyword descriptors). Table 1 describes the top-12 relevant laughter dimensions [22], shows the strongest positively and negatively associated descriptors, and shows the normalized dimensional LSA weight.

Table 1: Description of the top-12 LSA dimensions of female laughter [22]. A description is given, with the strongest above-threshold positively and negatively associated keywords. The 3rd column gives the normalized LSA dimensional weight. The top 12 dimensions accumulate about 25% of the model variance.

Dim	Laughter Dimension (LSA) [22]	Wt
L1	**Opposing qualities which vary widely**	**.053**
Neg:	*Low, happy, fast, slow, scared, & many others*	
L2	**Genuine happiness; sustained, voiced giggles**	**.029**
Pos:	Happy, genuine, giggle, chuckle, long	
Neg:	*Scared, air, gasp, breathy, quiet, short, soft*	
L3	**Short, sad, low-pitched, voiced chuckles**	**.022**
Pos:	Short, chuckle, low	
Neg:	*Happy, giggle, long, inhale, exhale, gasp*	
L4	**Fast, sure, simultaneous talking & laughing**	**.018**
Pos:	Fast, feminine, talking	
Neg:	*Surprised, nervous*	
L5	**Deep, resonant, and slow**	**.018**
Pos:	Sincere, deep, resonant, slow, relaxed	
Neg:	*Nervous, surprised*	
L6	**Soft, fast, and gruff**	**.018**
Pos:	Quiet	
Neg:	*Feminine, slow*	
L7	**Gentle, quiet, sustained, and nervous**	**.017**
Pos:	Nervous, worried, quiet, soft	
Neg:	*Surprised, short, loud*	
L8	**Surprised and shocked**	**.017**
Pos:	Surprised, shocked, alarmed	
Neg:	*Happy, sad*	
L9	**Nervous, unsure, tense, amusement**	**.017**
Pos:	Quiet, amused, nervous, unsure	
Neg:	*Soft*	
L10	**Sustained, nervous, fast, and voiced**	**.016**
Pos:	Nervous, fast, long	
Neg:	*Airy*	
L11	**Loud, strong, syllables**	**.016**
Pos:	Huh	
Neg:	*Quiet, feminine*	
L12	**Sarcastic and confident**	**.015**
Pos:	Sarcastic, sure	
Neg:	*Surprised*	

3.2. Perception of General Expressivity

In a previous study [24], we extracted expressive utterances (phrases and short sentences) from the same female talkers, and asked Mechanical Turk workers to provide three or more keywords describing the general expressivity in the voice. We performed LSA across these

Authorized licensed use limited to: University of Illinois. Downloaded on April 27,2022 at 15:50:03 UTC from IEEE Xplore. Restrictions apply.

clips and keywords, and used similar methods to discover a set of general expressive dimensions. These utterances contained many of the laughter segments which we examined for laughter perception in section 3.1. Table 2 shows the top-4 general, high-level expressive dimensions previously discovered in the corpus, and the number of laughter clips which are strongly associated with each of these dimensions via the context of the utterance.

Table 2: Description of the top-4 general, high-level expressive dimensions [24] in the corpus. The last column shows the number of strongly-associated laughter clips in each dimension. These dimensions were selected for analysis here because they were the strongest dimensions from the LSA analysis, and they contained a sufficient amount of laughter (some dimensions of general expression did not contain any laughter).

Dim	High-level Expressive Dimension (LSA) [24]	#
G1	Sincere, high-energy/high-affect, with laughter	15
G2	Joking, sarcastic, nervous speech, with laughter	18
G3	Low affect, with nervous energy	10
G4	Positive affect, with reflection and calm	8

The next section presents the feature discovery process for laughter and resulting laughter vector model. Then, it shows how the laughter vector can be used in the modeling of generalized vocal expression in a corpus. In this way, perception-grounded laughter models can be used in the modeling of higher-level perception-grounded models of general expressivity, and in some cases, can improve the performance of these higher-level models of expression.

4. DIMENSIONAL MODELING

RQ2 is answered by analyzing the discovered dimensions of laughter, and building acoustic regression models which produce a laughter vector when applied to laughter audio. Next, RQ3 is addressed by using the laughter vector outputs as inputs to higher-level expressive dimension classifiers. In this way, we show how laughter episodes embedded in context of containing phrases can be used to identify general modes of expressivity for these phrases.

4.1. Modeling Dimensions of Laughter

To predict the degree of association of each laughter clip with each of the discovered dimensions of laughter, we leverage the LSA model. The values in the LSA projection matrix (laughter clips projected onto the 12 dimensions) provide the training values for the regression models for each dimension of laughter.

The baseline feature set included both the laughter types described in section 3, and the openSMILE ComParE13 feature set [5] (60 msec frames and a 10-msec hop). We selected the ComParE13 for its wide range of acoustic features and successful application in the Interspeech Paralingual Challenges. Since this feature set was so large, we removed all low-variance features with $\sigma^2<0.05$ from consideration, and then performed feature ranking and selection within training folds, using 5-fold cross validation with p=0.01. Table 3 summarizes the best results for each laughter dimension regression model, and presents 1) the best 5 feature groups overall, and 2) the best R, mean squared error, and number of features retained by ranking. Laughter dimensions L1, L8, L9, and L11 were removed from consideration because they did not have a sufficient number of representative laughter examples.

Table 3: Ridge regression performance for each viable dimension of laughter. The top 5 feature groups are shown here, with a '*' indicating multiple statistical variants (e.g., skewness, kurtosis, percentile, etc.) on the base feature. The third column shows the Spearman R, the mean squared error, and the number of features retained in the final "production" model.

Dim	Best 5 Feature Groups (* indicates multiple statistical functionals)	R *mse* (#)
L2	pcm_fftMag_spectralFlux_sma* audSpec_Rfilt_sma[6, 18, 22]* pcm_RMSenergy_sma* audspecRasta_length_L1norm* mfcc_sma_de[4]*	**0.65** *0.028* (60)
L3	pcm_fftMag_fband1000-4000_sma_quartile1 mfcc_sma[2]* F0final_sma* pcm_fftMag_spectralRollOff* Joint_Laughter	**0.60** *0.001* (37)
L4	pcm_fftMag_spectralSlope_sma_de* pcm_RMSenergy_sma_upleveltime25 pcm_fftMag_fband250-650_sma_flatness pcm_fftMag_spectralSlope_sma* pcm_fftMag_spectralHarmonicity_sma*	**0.42** *0.001* (11)
L5	mfcc_sma_de[8]_maxSegLen mfcc_sma[4]_maxSegLen audspec_Rfilt_sma[21]_kurtosis mfcc_sma_de[14]_upleveltime25 audSpec_Rfilt_sma[24]_skewness	**0.28** *0.023* (10)
L6	audSpec_Rfilt_sma[10, 19, 23, 24, 25]* audSpec_Rfilt_sma_de[6, 24, 25]* pcm_fftMag_spectralVariance_sma* jitterLocal_sma_minPos F0final_sma_quartile2	**0.16** *0.020* (22)
L7	pcm_RMSenergy_sma* logHNR_sma_de_upleveltime25 pcm_fftMag_spectralHarmonicity_sma* Joint_Speak Audspec_lengthL1norm_sma*	**0.41** *0.021* (26)
L10	audSpec_Rfilt_sma[24]_upleveltime50 audSpec_Rfilt_sma[6]_minPos mfcc_sma[9]_peakRangeRel audSpec_Rfilt_sma_de jitterDDP_sma_lpc4	**0.11** *0.028* (9)
L12	pcm_fftMag_spectralCentroid_sma* mfcc_sma[10]_maxPos audSpec_Rfilt_sma_de[17] pcm_zcr_sma_upleveltime50 audSpec_Rfilt_sma_de[2]_kurtosis	**0.13** *0.038* (7)

Authorized licensed use limited to: University of Illinois. Downloaded on April 27,2022 at 15:50:03 UTC from IEEE Xplore. Restrictions apply.

Several of the selected features emphasize spectral content below the range of normal female speaking voice. These very low frequencies (e.g., mfcc_sma[2], mfcc_sma_de[4], and audSpec_Rfilt_sma_de[2] in L2, L3, and L12) may be reflecting laughter periodicity. This suggests that sad (L3) or sarcastic (L12) laughter pulses at a slower rate than laughter resulting from sincere happiness (L2). It is also interesting to see that sad laughter (L3) may frequently be shared between conversation partners (Joint_Laughter feature), and that nervous laughter (L7) occurs when one party is laughing, and the other is speaking (Joint_Speak feature). The social element of laughter, therefore, has an important part in defining the expressive dimensionality of laughter.

The 250-650 Hz band is important to distinguishing modal speech [23], which may be the reason it is a distinguishing feature for L4, which contains simultaneous laughing and talking. Frequency instability, or jitter, may be expected in a nervous person's voice (L10), or a gruff person's voice (L6), especially if the gruffness comes from vocal quality changes, such as creakiness. Many of the important features overall are RASTA-filtered segments of the spectrum (Rfilt), and RASTA suppresses spectral components that vary at rates different from the typical rate of change in speech. Also, pcm_RMSEnergy_sma features capture characteristics of the signal frame energy (related to the loudness contour), which varies with laugher pulses, and varies greatly in quality across different kinds of laughter. pcm_fftMag_spectralHarmonicity_sma measures the quality of the harmonics in the signal, which would be distinctly different in joint speaking and laughing (L7).

4.2. Modeling General Expressivity Using Laughter

Table 4: Using laughter segments to classify the expressive quality of the containing phrase. The first column identifies the high-level, expressive LSA dimension of the containing phrase (from Table 2). The best components of the laughter vector are given for each classifier in column 2 (laughter features correspond to the LSA dimensions in Tables 1 and 3), and the Average Unweighted Recall of the laughter-only classifiers is given in column 3. The last column references the results of using acoustic features to classify the dimension directly, from a previous study [24].

Dim	Best Laughter Features	AUR Laughter	AUR Acoustic[24]
G1	L4	**0.67**	**0.79**
G2	ALL	**0.67**	**0.60**
G3	L3, L4, L5	**0.71**	**0.80**
G4	L5, L12	**0.75**	**0.61**

A laughter segment can be used to predict the general mode of expressivity within its phrase or sentence context. Passing each laughter clip through each of the laughter regression models produces a laughter vector, which in turn, is used to train classifiers to recognize the higher-level modes of expressivity described in Table 2. Table 4 compares the performance of linear SVM classifiers which used only the laughter vector to predict the expressive dimension of the containing speech segment, versus the performance of classifiers which used the acoustic features from the speech segment to predict the expressive dimension. The results suggest that examining laughter segments contained in spoken phrases can improve recognition of some difficult-to-recognize modes of human expression, such as sarcasm and humor, without having to resort to examining the language text. This can be helpful in situations where text is unavailable, or expensive to acquire.

5. DISCUSSION AND CONCLUSIONS

This paper presented a technique for discovering perception-grounded dimensions of laughter in a corpus and creating acoustic models of the resulting laughter dimensions, which when applied to laughter samples, produce a laughter vector. This process directly addresses RQ1 and RQ2, and provides the necessary input to address RQ3. We have shown that classification of general vocal expression dimensions can be improved by using the laughter vector, and that the laughter vector can be helpful in classifying some modes of expressivity which are known to be difficult to classify using acoustics only, such as sarcasm.

This technique enables detailed analysis of the relationships among voice quality, nonverbal quality, emotion, and prosody by examining the co-occurrences of descriptors within dimensions, which could potentially improve emotion recognition in the future. As we have demonstrated here, it also allows the leverage of one set of discovered dimensions in the recognition of another. The technique scales and its models can be hierarchical. Low-level acoustic features can in that way be associated, with a quantifiable weight, to an arbitrary number of expressive dimensions. The technique could be applied to a wide range of domains, not just speech. Finally, the dimensions discovered are anchored in human perception, which tends to encourage production of software components aligned with what people see and hear. The results, therefore, more naturally encourage human-friendly application development and interface design.

Future work could expand the data set to include a greater number and range of speakers from both genders, exploration of gender-specific and gender-neutral models, transfer learning explorations, and a wider range of speaking styles. This expanded data set could potentially enable analysis using a wider range of techniques.

6. REFERENCES

[1] Gouzhen An, David Guy Brizan, and Andrew Rosenberg, "Detecting Laughter and Filled Pauses Using Syllable-based Features," INTERSPEECH 2014.

Authorized licensed use limited to: University of Illinois. Downloaded on April 27,2022 at 15:50:03 UTC from IEEE Xplore. Restrictions apply.

[2] Anton Batliner, Stefan Steidl, Florian Eyben, and Bjorn Schuller, "On Laughter and Speech-Laugh, Based on Observations of Child-Robot Interaction," In Jurgen Trouvain and Nick Campbell (eds), The Phonetics of Laughing, Sarland University Press, January 2011.

[3] Gregory A. Bryant et al., "The Perception of Spontaneous and Volitional Laughter Across 21 Societies," Association for Psychological Science, 29(9): 1515-1525, 2018.

[4] Sri Harsha Dumpala, Ashish Panda, and Sunil Kumar Kopparapu, "Analysis of the Effect of Speech-Laugh on Speaker Recognition System," Interspeech 2018.

[5] Florian Eyben, Martin Wollmer, and Bjorn Schuller, "openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor," MM 2010.

[6] Gabor Gosztolya, "Optimized Time Series Filters for Detecting Laughter and Filler Events," Interspeech 2017

[7] Rahul Gupta, Theodora Chaspari, Panayiotis Georgiou, David Atkins, and Shrikanth Narayanan, "Analysis and modeling of the role of laughter in Motivational Interviewing based on psychotherapy conversations," Interspeech 2015.

[8] Rahul Gupta, Panayoitis G Georgiou, David C. Atkins, Shrikanth S. Narayanan, "Predicting client's inclination towards target behavior change in notivational interviewing and investigating the role of laughter," Interspeech 2014

[9] Gerhard Hagerer, Nicholas Cummins, Florian Eyben, and Bjorn Schuller, "Did you laugh enough today?" Deep Neural Networks for Mobile and Wearable Laughter Trackers, Interspeech 2017

[10] Lakshmish Kausik, Abhijeet Sangwan, and John H.L. Hansen, "Laughter and Filler Detection in Naturalistic Audio," Interspeech 2015.

[11] Lyndon S. Kennedy and Daniel P.W. Ellis, "Laughter Detection in Meetings," in Proc. NIST ICASSP 2004 Meeting Recognition Workshop, Montreal, Canada, pp. 118-121, May 2004.

[12] Mary Tai Knox and Nikki Mirghafori, "Automatic Laughter Detection Using Neural Networks," Interspeech 2007.

[13] Teun F. Krikke, Khiet P. Truong, "Detection of nonverbal vocalizations using Gaussian Mixture Models: looking for fillers and laughter in conversational speech, " Interspeech 2013.

[14] Thomas K. Landauer, Peter W. Foltz, and Darrell Laham, "An Introduction to Latent Semantic Analysis," Discourse Processes, 25(2&3):259-284, 1998.

[15] Library of Congress Veterans History Project, available at https://www.loc.gov/vets/, accessed 10/20/18.

[16] Willem A. Melder, Khiet P. Truong, Marten Den Uyl, David A. Van Leeuwen, Mark A. Neerincs, Lodewijk R. Loos, and B. Stock Plum, "Affective multimodal mirror: sensing and eliciting laughter," HCM 2007.

[17] C. Menezes and Y. Igarashi, "The speech laugh spectrum," in Proc. 6th International Seminar on Speech Production ISSP, Dec. 2006.

[18] Vinay Kumar Mittal, B. Yegnanarayana, "Study of changes in glottal vibration characteristics during laughter," Interspeech 2014.

[19] Jieun Oh, Eunjoon Cho, and Malcolm Slaney, "Characteristic Contours of Syllabic-level Units in Laughter," Interspeech 2013.

[20] Jieun Oh, Ge Wang, "Laughter modulation: from speech to speech-laugh," Interspeech 2014.

[21] Andrew Ortony and Terence J. Turner, "What's Basic About Bsaic Emotions?" Psychological Review, 97(3): 315-331, 1990.

[22] Mary Pietrowicz, "Exposing the Hidden Vocal Channel: Analysis of Vocal Expression," Dissertation, University of Illinois at Urbana-Champaign, 2017.

[23] Mary Pietrowicz, Mark Hasegawa-Johnson, and Karrie Karahalios, "Acoustic Correlates for perceived effort levels in male and female acted voices," Journal of the Acoustical Society of America, 142(2):792, 2017.

[24] Mary Pietrowicz, Mark Hasegawa-Johnson, and Karrie Karahalios, "Discovering Dimensions of Perceived Vocal Expression in Semi-Structured, Unscripted Oral History Accounts," ICASSP 2017.

[25] Sophie Scott, Nadine Lavan, Sinead Chen, and Carolyn McGettigan, "The social life of laughter," Trends Cog Sci 18(12):618-629, 2014.

[26] Diana P. Szameitat, Benjamin Kreifelts, Kai Alter, Andre J. Szameitat, Anette Sterr, Wolfgang Grodd, and Dirk Wildgruber, "It is not always tickling: Distinct cerebral responses during perception of different laughter types," NeuroImage 53: 1264-1271, 2010.

[27] Khiet P. Truong and Jurgen Trouvain, "On the acoustics of overlapping laughter in conversational speech," Interspeech 2012.

[28] Khiet P. Truong, and Jurgen Trouvain," Investigating prosodic relations between initiating and responding laughs," Interspeech 2014.

[29] Bekir Berker Turker, Zana Bucinca, Engin Erzin, Yucel Yernez, Metin Sezgin, "Analysis of Engagement and User Experience with a Laughter Responsive Social Robot," Interspeech 2017.

Authorized licensed use limited to: University of Illinois. Downloaded on April 27,2022 at 15:50:03 UTC from IEEE Xplore. Restrictions apply.