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 Thomas J Watson Research Center[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.

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?

Data Corpus

The study corpus included selected oral history interviews from the Veterans History Project at the Library of Congress[1]. These interviews had a similar format across all samples, covered similar topics of discussion across speakers, encouraged unscripted storytelling and discussion, and provided a diverse range of natural laughter to explore.

- **Library of Congress Veterans' History Project**
- **10 representative female speakers**
- **120 laughter events**
- **Young veterans from recent conflicts, post-1995**
- **Original recordings made on modern media**
- **Interviews were approximately 1 hour in duration**

Prior work[2] evaluated the overall expressive dimensionality of the surrounding speech context (sentences and phrases) of the laughter events.

Table 1: ***Perceived expression in the speech phrases and sentences which surrounded the laughter.***

Dim[2]	Description
G1	Sincere, high-energy/high-affect, with laughter
G2	Joking, sarcastic, and nervous, with laughter
G3	Low affect, with nervous energy
G4	Positive affect, with reflection and calm

How would listeners perceive the laughter alone, and could analysis of laughter help us recognize these higher-level modes of expression?

Perception Study

Figure 1: 1200 Mechanical Turk workers described 120 representative laughter samples using 4000+ words. Over half of the words described a wide range of nuanced emotion in the laughter. Almost 40% of the words described a narrow range of voice quality and prosody in the laughter.

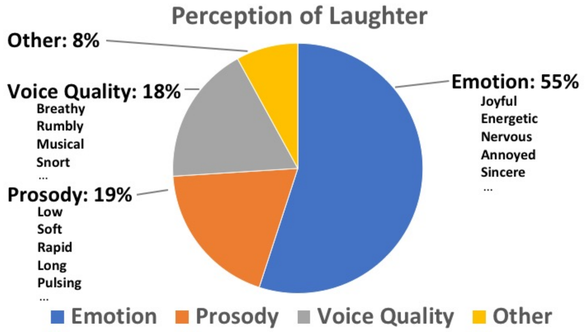

Table 2: ***Latent semantic analysis techniques applied to the descriptors and sound clips revealed the types of expressive laughter present in the corpus***. Positive and negative descriptor associations tell us what each laughter dimension is and isn't.

Dim	Laughter Dimension Description (LSA)	Weight
L1	**High-variance laughter with opposing qualities**	0.053
Neg:	*Low, happy, fast, slow, scared, & many others*	
L2	**Genuine happiness; sustained, voiced giggles**	0.029
Pos:	Happy, genuine, giggle, chuckle, long	
Neg:	*Scared, air, gasp, breathy, quiet, short, soft*	
L3	**Short, sad, low-pitched, voiced chuckles**	0.022
Pos:	Short, chuckle, low	
Neg:	*Happy, giggle, long, inhale, exhale, gasp*	
L4	**Fast, sure, simultaneous talking and laughing**	0.018
Pos:	Fast, feminine, talking	
Neg:	*Surprised, nervous*	
L5	**Deep, resonant, and slow**	0.018
Pos:	Sincere, deep, resonant, slow, relaxed	
Neg:	*Nervous, surprised*	
L6	**Soft, fast, and gruff**	0.018
Pos:	Quiet	
Neg:	*Feminine, slow*	
L7	**Gentle, quiet, sustained, and nervous**	0.017
Pos:	Nervous, worried, quiet, soft	
Neg:	*Surprised, short, loud*	
L8	**Surprised and shocked**	0.017
Pos:	Surprised, shocked, alarmed	
Neg:	*Happy, sad*	
L9	**Nervous, unsure, tense, amusement**	0.017
Pos:	Quiet, amused, nervous, unsure	
Neg:	*Soft*	
L10	**Sustained, nervous, fast, and voiced**	0.016
Pos:	Nervous, fast, long	
Neg:	*Airy*	
L11	**Loud, strong, syllables**	0.016
Pos:	Huh	
Neg:	*Quiet, feminine*	
L12	**Sarcastic and confident**	0.016
Pos:	Sarcastic, sure	
Neg:	*Surprised*	

Laughter Models

To address **RQ2**, we created an ensemble of regression models which measured how closely a laughter sample matched the top 8 laughter types we discovered. Features were selected from an extended openSMILE[3] ComParE13 set. Results are sonic **laughter fingerprints** via a **LaughterToVec** style model.

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, percentile, etc.) on the base feature. The 3rd column shows the Spearman R, the mean squared error (mse), and the number of features (#) retained in the final model.

Dim	Best 5 Feature Groups *multiple statistical functionals	R *mse* (#)
L2	**Genuine happiness; sustained, voiced giggles** 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	**Short, sad, low-pitched, voiced chuckles** pcm_fftMag_fband1000-4000_sma_quartile1 mfcc_sma[2]* F0final_sma* pcm_fftMag_spectralRollOff* Joint_Laughter	**0.60** *0.001* **(37)**
L4	**Fast, sure, simultaneous talking and Laughing** 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	**Deep, resonant, and slow** 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	**Soft, fast, and gruff** 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* **(26)**
L7	**Gentle, quiet, sustained, and nervous** pcm_RMSenergy_sma* logHNR_sma_de_upleveltime25 pcm_fftMag_spectralHarmonicity_sma* Joint_Speak Audspec_lengthL1norm_sma*	**0.41** *0.021* **(26)**
L10	**Sustained, nervous, fast, and voiced** 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	**Sarcastic and confident** 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)**

Using Laughter

Can laughter fingerprints could help us identify general modes of vocal expression **(RQ3)**? The answer is yes for some difficult-to-recognize kinds of vocal expression, notably sarcasm.

Table 4: ***Using laughter segments to classify the expressive quality of the containing phrase vs. using acoustic features alone.*** The first column identifies the high-level expressive dimension of the containing phrase. The second column shows the best components of the laughter fingerprint for recognizing the corresponding General Expressive Dimension. The 3rd and 4th columns show the Average Unweighted Recall (AUR) of classifiers which use only laughter fingerprints as features, and the AUR of classifiers which use only low-level acoustic features.

General Expressive Dimension[2]	Best Laughter Features	AUR Laughter Only	AUR Acoustic Only
G1: Sincere, high-energy, high-affect	L4	0.67	0.79
G2: Joking, sarcastic, & nervous	ALL	0.67	0.60
G3: Low affect, with nervous energy	L3, L4, L5	0.71	0.80
G4: Positive affect, with reflection and calm	L5, L12	0.75	0.61

Conclusions

This work makes the following contributions:

1. **A new technique for discovering perception-grounded dimensions of laughter in a corpus (RQ1)**
2. **A new technique for analyzing laughter as sonic laughter fingerprints via a LaughterToVec style model which measures how closely a given laugh matches the characteristics of the discovered dimensions of laughter (RQ2)**
3. **Demonstration of the use of "laughter fingerprints" to improve the recognition of sarcasm and other difficult-to-recognize expression in speech (RQ3)**

Future work will expand the exploration of laughter, apply the techniques to other kinds of paralingual expression, and explore the relationship of laughter to health and wellness states.

References

1. Veterans History Project at the Library of Congress, available at https://www.loc.gov/vets/
2. Mary Pietrowicz, Mark Hasegawa-Johnson, and Karrie Karahalios, "Discovering Dimensions of Perceived Vocal Expression in Semi-structured, Unscripted Oral History Accounts," ICASSP 2017.
3. OpenSMILE feature extraction tool, available at https://www.audeering.com/opensmile/