Whom we laugh with affects how we laugh

Nick Campbell

NiCT/ATR-SLC
National Institute of Information and Communications Technology
& ATR Spoken Language Communication Research Labs
Keihanna Science City, Kyoto 619-0288, Japan
nick@nict.go.jp, nick@atr.jp

ABSTRACT

This paper describes work that shows how the acoustic features of laughter in Japanese speech vary according to conversational partner, reflecting the social status of laughter, and confirming that even such a simple sound is affected by non-linguistic factors such as social or intercultural relationships. Neural networks were successfully trained to identify the nature of the interlocutor from principal components of the acoustic and prosodic features of the laughing speech.

Keywords: Laughter, laughing speech, voice quality, acoustic characteristics, principal-component analysis, neural-network training

1. INTRODUCTION

The acoustics of laughter have been shown to be both highly complex and highly variable [1] with voiced and unvoiced variants functioning separately and having different effects [2]. However, most studies of laughter have been concerned with reactions to media rather than with laughter in interactive conversational situations. Recent work by [3] has shown laughter in conversation to be much more frequent than has been described previously in the literature, and suggests that this form of interactive laughter may primarily serve both to regulate the flow of the interaction and to mitigate the meaning of a preceding utterance. High intra-individual variability which greatly exceeded the parameter variability between subjects was found in the acoustic parameters of this type of laughter. The present paper extends this work to examine how laughter in the speech of two Japanese adults also varies systematically according to the nature of the interlocutor.

In this paper we make use of a global measure of the acoustics of laughter, derived from a principal component analysis of fourteen basic measures of prosodic and spectral characteristics incorporating voice quality [4]. It has been shown elsewhere [5] that this measure correlates closely with the changes in speaking style that occur with differences in familiarity between a speaker and a listener, and with differences in the ease of conversation that arise from e.g., cross-cultural or cross-language interactions. In the present paper we examine the changes in these characteristics that occur in the laughter and laughing speech of two Japanese individuals, one man and one woman, in conversations with four strangers over a period of time. The speech is in Japanese, but it is likely that the phonetic and prosodic characteristics of laughter are common to all people of whatever language background. However, the nature and style of laughing may of course vary considerably according to cultural and situational constraints.

Table 1: Counts of utterances extracted from the corpus. All are laughs, those on the right are laughing while speaking. C and E represent Chinese and English native-language partners, F and M the sex of the interlocutor. The sex and language of the speaker is shown in the second row

	laughs		+ speech	
	JF	JM	JF	JM
CF	201	241	131	214
CM	174	174	93	156
EF	350	401	140	173
EM	228	232	100	122

2. DATA

The speech data were recorded over a period of several months, with paid volunteers coming to an office building in a large city in Western Japan once a week to talk with specific partners in a separate part of the same building over an office telephone. While talking, they wore a head-mounted Sennheiser HMD-410 close-talking dynamic microphone and recorded their speech directly to DAT (digital audio tape) at a sampling rate of 48kHz. They did not see their partners or socialise with them outside of the recording sessions. Partner combinations were controlled for sex, age, and familiarity,

and all recordings were transcribed and time-aligned for subsequent analysis. Recordings continued for a maximum of ten sessions between each pair. Each conversation lasted for a period of thirty minutes.

In all, ten people took part as speakers in the corpus recordings, five male and five female. Six were Japanese, two Chinese, and two native speakers of American English. All conversations were held in Japanese. There were no constraints on the content of the conversations other than that they should occupy the full thirty-minute time slot. Partners were initially strangers to each other, but became friends over the period of the recordings. The conversations between the three pairs of Japanese speakers form the main part of this corpus [5], and the conversations with non-native speakers form a subpart which is reported here. The non-native speakers were living and working in Japan, competent in Japanese, but not at a level approaching native-speaker fluency.

The speech data were transferred to a computer and segmented into separate files, each containing a single utterance. Laughs were marked with a special diacritic, and laughing speech was also bracketed to show by use of the diacritic which sections were spoken with a laughing voice. Laughs were transcribed using the Japanese Katakana orthography, wherever possible, alongside the use of the symbol.

The present analysis focusses on these two types of laughter as produced by the two Japanese speakers who spoke to the highest number of partners (see Table 1 for counts), and examines the changes depending on relationship with the interlocutor as characterised by native-language and sex.

3. MODELLING THE LAUGHS

Laughter was very common in the speech of all the conversation participants. Their situation was unusual in that although they did not initially know each other, they were required to talk over a telephone line (with no face-to-face contact) for a period of thirty minutes each week for five weeks. They were all paid and willing volunteers and knew that their recordings would be used for telecommunications research, but they had no detailed knowledge about the purpose of the recordings. Over the period of five conversations, they came to know each other quite well.

The transcribed speech files containing laughter were processed by a computer program to extract a set of acoustic features for each utterance. Since the utterances were typically short, we used a single value for each feature to describe an utterance. The features included pitch, power, duration, and spectral shape. Pitch was described by the mean, maximum, minimum, location of the peak in the utterance, and degree of voicing throughout the utterance. Power was described by the mean, maximum, minimum, and location of the peak in the utterance. Duration of the whole utterance was expressed as a log value, and a simple estimate of speaking rate was made by dividing the duration by the number of moraic units in the transcription. Spectral shape was described by the location and energy of the first two harmonics, the amplitude of the third formant, and the difference in energy between the first harmonic and the third formant (h1-a3, proposed by Hansen as the best measure for describing breathiness in her study of the voice quality of female speakers [6]). All these measures were produced automatically using the Tcl/Tk "Snack" audio processing library [7]. Thus for each laughing utterance in the conversations, we produced a vector of values corresponding to its acoustic characteristics.

Table 2: Cumulative proportion of the variance accounted for by the principal component analysis. 'f' and 'm' stand for female and male, and 'l' and 's' for laughter and laughing-speech respectively. Only the first 10 components are shown

	pc1	pc2	pc3	pc4	pc5	pc6	pc7	pc8	pc9	pc10
f-l	.23	.43	.54	.64	.72	.78	.84	.88	.92	.95
m-l	.31	.45	.57	.66	.74	.79	.84	.89	.92	.95
f-s	.21	.35	.49	.58	.65	.72	.79	.85	.89	.93
m-s	.19	.33	.46	.55	.64	.72	.78	.84	.89	.93

3.1. Principal Component Analysis

To simplify the use of these acoustic features in training a statistical model, we performed a principal component analysis [8] using the "princomp" function call in R [9]. The first three principal components account for about 50% of the variance in the acoustic data, and the first seven components together account for more than 80%. Table 2 shows that the first five principal components accounted for approximately 73% of the acoustic and prosodic variance in the laughs, and approximately 65% of the acoustic and prosodic variance in the laughing speech. The limited phonetic component of simple laughter makes it acoustically less variable than the laughing speech, and hence slightly easier to model.

3.2. Neural Network Training

In order to determine whether the variance observed in these laughs was related in any way to the nature of the interlocutor, a neural network was trained to learn the mapping between the first five (5) principal components and a label representing either (a) Chinese vs. English, or (b) male vs. female.

Table 3: Raw scores for the neural network trained to distinguish between Chinese and English-speaking partners. Here, C stands for Chinese, and E for English, with X for indeterminate ('don't know') predictions.

laughs						
	JF			JM		
→	E	C	X	E	C	X
CF	34	304	162	40	338	122
CM	41	299	160	33	350	117
EF	326	47	127	318	56	126
EM	284	45	171	350	30	120
laughing speech						
	JF			JM		
→	E	C	X	E	C	X
CF	53	353	94	34	369	97
CM	39	383	78	49	351	100
EF	369	42	89	358	45	97
EM	388	38	74	363	37	100

Table 4: Raw scores for the neural network trained to distinguish between male and female partners. Here, M stands for male, and F for female, with X for indeterminate predictions. In all cases, the 'correct' answer predominates.

laughs						
	JF			JM		
→	M	F	X	M	F	X
CF	71	259	170	34	337	129
CM	271	24	205	318	59	123
EF	14	352	134	26	349	125
EM	277	22	201	326	53	121
laughing speech						
	JF			JM		
→	M	F	X	M	F	X
CF	39	333	128	37	341	122
CM	352	26	122	358	43	99
EF	43	324	133	46	329	125
EM	332	40	128	353	27	120

A back-propagation neural network was constructed with five input neurons representing the activation of the first five (5) principal components of the acoustics of the laughter, or laughing speech, with a layer of seven (7) intermediate neurons and an output layer of two (2) neurons representing either male or female partners, or Chinese native-language or English native-language partners depending on the training session. The $nnet$ function of R was used for this with the following arguments:
$pcnet = nnet.formula(who \sim pc1 + pc2 + pc3 + pc4 + pc5, size = 7, rang = 0.1, decay = 5e-4, maxit = 500, trace = F)$
and repeatedly trained for each combination of speaker, laughing type, and interlocutor pattern.

We randomly selected from the utterances shown in Table 1 a subset of fifty (50) tokens for each partner of male and female laughter and laughing speech samples for training (giving 4×200 tokens in all) and a separate set of 50 each for testing in each category. Using an arbitrary threshold, values greater than 0.5 in the output neurons were taken as positive, less than -0.5 as negative, and values between -0.5 and 0.5 were taken to indicate that the network could not distinguish between training classes on the basis of the five principal component values for each token.

The network was trained with fifty (50) samples each of (a) laughter and (b) laughing speech randomly selected from conversations with each class of partner (c,d), giving a training vector of two-hundred ($4 \times 50 = 200$) samples. The trained network was then tested on a completely different vector of two-hundred (200) samples from a different random selection under the same criteria.

Because the networks are randomly initialised, and can produce different results with each training session, we performed ten (10) training and testing cycles for each combination and summed the results for each prediction category. These are the figures reported in the Tables. Tables 3 and 4 give the raw training results for each combination. The labels 'E', 'C', and 'X' in Table 3 indicate predictions for English, Chinese and 'don't-know' for Chinese female partner (CF), Chinese male partner (CM) etc. It can be seen from the tables that the networks successfully identify the partner from the acoustics of the laughter or laughing speech in the majority of cases.

4. RESULTS

Tables 5 and 6 show expanded summaries of the data in Tables 3 and 4 for a comparison of differences between the various prediction tasks. Statistics for the networks trained to detect the sex of the partner from the rotated acoustic parameters are shown in Table 5, and those for the Chinese/English discrimination in Table 6. The two leftmost columns in the tables provide summed results, disregarding individual partner differences (which can be examined from Tables 3 or 4). No test is necessary to see that these differences are significant, with more than six hundred correct responses against less than a hundred false responses in every case.

The centre two columns of the table are more revealing. They show counts of hits, misses, and

Table 5: Summed scores for the trained networks predicting male/female partner distinction from acoustic parameters. Indeterminate prediction results are shown in brackets, see text for an explanation.

laughs - JF					
discrim.		accuracy		JF	
85	611	131	1159	279	2500
548	46	-	(710)	-	(1221)
laughing speech - JF					
discrim.		accuracy			
82	657	148	1341	-	-
684	66	-	(511)	-	-
laughs - JM					
discrim.		accuracy		JM	
60	686	172	1330	325	2711
644	112	-	(498)	-	(964)
laughing speech - JM					
discrim.		accuracy			
83	670	153	1381	-	-
711	70	-	(466)	-	-

Table 6: Summed scores for the trained networks predicting Chinese/English partner distinction from acoustic parameters. Indeterminate prediction results are shown in brackets, see text for an explanation.

laughs - JF					
discrim.		accuracy		JF	
75	603	167	1213	339	2706
610	92	-	(620)	-	(955)
laughing speech - JF					
discrim.		accuracy			
92	736	172	1493	-	-
757	80	-	(335)		
laughs - JM					
discrim.		accuracy		JM	
73	688	159	1356	324	2797
668	86	-	(485)	-	(879)
laughing speech - JM					
discrim.		accuracy			
83	720	165	1441	-	-
721	82	-	(394)	-	-

'don't-know' responses for each class of speaker and laughing style. The two columns on the right of the table summarise these figures across each style of laughter to provide overall scores for each speaker.

Pearson's Chi-Square test [10] was used to compare each pair of results, and only JF(m/f) and JM(m/f) showed any significant differences.
JF-M/F accuracy (85, 611, 46, 548):
$\rightarrow \chi^2$ = 6.53, df = 1, p = 0.01059 (signif)
JF-M/F confidence (611, 304, 548, 406):
$\rightarrow \chi^2$ = 16.87, df = 1, p = 3.987e-05 (signif)
JM-M/F accuracy (60, 686, 112, 644):
$\rightarrow \chi^2$ = 16.32, df = 1, p = 5.349e-05 (signif)
JM-M/F confidence (686, 254, 644, 244):
$\rightarrow \chi^2$ = 0.02, df = 1, p = 0.8678 (n.s.).

No other differences between correct and false partner discriminations are significant. However, the difference in performance for JF overall, comparing discrimination success, is considerable:
JF-M/F overall (964, 2711, 1221, 2500):
$\rightarrow \chi^2$ = 38.17, df = 1, p = 6.478e-10 (signif)
cf JM-M/F overall (955, 2706, 879, 2797):
$\rightarrow \chi^2$ = 4.51, df = 1, p = 0.03373 (n.s.).

However, even for the least successful case (predicting the sex of the interlocutor from the style of JF's laughter) the network achieves 62.5% accuracy against a chance score of 50%. The male speaker's laughing idiosyncrasies allow the network to predict the sex of his interlocutor at 67.7% accuracy. The female speaker, differentiates her style of laughter when talking with foreigners sufficiently for the network to discriminate at 67.5%, and the male at an even higher rate of 70%.

From these stringent training and testing conditions one can conclude that the network is indeed able to generalise from the features of the acoustics in order to be able to identify the interlocutor at rates significantly better than chance. This confirms that speakers modify their laughter in a consistent way that indicates something about the nature of their relationship with the interlocutor.

5. DISCUSSION

JM laughs most with CF and EF; JF laughs least with CM and EM. Both laugh much more with EF (whose Japanese is less than fluent). It remains as future work to examine the nature of those relationships and the role of the individual acoustic features in triggering the different perceptions. However, some details of the acoustic mapping are given in Table 7 which shows first three principal components in each situation. The numbers are related to the strength of contribution of each acoustic feature in each component. For simplicity, values lower than 25 have been replaced by dashes to facilitate comparison. The table shows that in all cases the breathiness of the voice, as indicated by h1-a3 (a measure of spectral tilt, derived from subtracting energy measured at the third formant from energy measured at the first harmonic) plays an important contribution with strong weightings in every case.

Table 7: Contribution values (rotations) of each prosodic or acoustic feature in the first three principal components of each speaker-laughing-style combination. Vertical bars separate the laughing styles, and values for pc1, pc2, pc3 are listed in order within each. Values less than 25 have been replaced by dashes for simplicity

	JF-laugh			JF l+sp			JM-laugh			JM-l+sp		
fmean	--	35	--	45	--	--	33	--	--	32	32	--
fmax	29	37	--	--	27	--	30	27	--	36	--	34
fmin	39	--	--	34	--	28	33	--	--	--	--	--
fpct	--	--	--	--	--	28	--	--	39	--	--	32
fvcd	26	41	--	--	35	--	26	37	--	33	26	--
pmean	--	46	--	--	51	35	30	39	--	39	46	--
pmax	37	--	--	--	46	--	38	--	--	36	33	--
pmin	--	36	25	--	--	42	--	49	--	--	30	--
ppct	19	--	36	28	--	33	--	30	--	--	--	--
h1h2	--	--	25	--	--	--	--	--	--	--	--	35
h1a3	40	--	44	41	27	28	32	--	48	32	40	38
h1	40	--	26	38	--	36	33	--	32	32	39	--
a3	--	--	45	--	--	--	--	--	49	--	--	45
dn	--	31	37	34	--	30	--	47	--	--	--	29

The speakers control their voices differently, both in simple laughter and in laughing speech, and the differences in pitch, loudness and tension of the voice, or breathiness, reveal characteristics related both to the sex of the interlocutor and to differences in cultural background.

6. CONCLUSION

This paper has described a brief study of laughs and laughing speech excised from the telephone conversations of two Japanese speakers talking with two male and two female partners. It presented results showing that the speakers vary their laughing styles according to the sex and nationality of the partner.

A neural network was trained to distinguish either the sex of the interlocutor or their social background, as characterised by native language, and differences in the success of the training were compared for each of these two dimensions and for each of the two speakers.

It was shown in previous work [11] that a speaker adapts her voice quality as well as speaking styles according to the nature of her relationship with the interlocutor. The present study provides additional evidence for this common-sense but largely unexplored phenomenon by showing that differences can be also be found in the types of laughter expressed by a further two male and female speakers of Japanese in telephone conversations with four partners each over a period of five weeks.

In separate work with a very large single-speaker corpus [12] we found that approximately one in ten utterances contains laughter. From among these laughing utterances, we were able to distinguish four types of laughter according to what each revealed about the speaker's affective state, and were able to recognise these different types automatically by use of Hidden Markov Models trained on laugh segments, with a success rate of 75%. In future work we will attempt a similar perceptual classification of the different types of laughter found in the present corpus, and will attempt to explain their interpretation in a social and discourse context.

Acknowledgement

This work is partly supported by the National Institute of Information and Communications Technology (NiCT), and includes contributions from the Japan Science & Technology Corporation (JST).

7. REFERENCES

[1] Bachorowski, J.-A., Smoski, M.J., & Owren, M.J. (2001). "The acoustic features of human laughter", Journal of the Acoustical Society of America, 110, pp.1581-1597.

[2] Bachorowski, J.-A., and Owren, M.J. (2001). "Not all laughs are alike; Voiced but not unvoiced laughter elicits positive affect in listeners", Psychological Scienece 12, pp.252-257.

[3] Vettin, J. & Todt, D. (2004): "Laughter in conversation: features of occurrence and acoustic structure". J. Nonverbal Behaviour. 28: 93-115.

[4] Ni Chasaide, A., and Gobl, C., (1997). "Voice source variation". In W. J. Hardcastle and J. Laver (Eds.), *The Handbook of Phonetic Sciences*, pp. 428-461. Oxford: Blackwells.

[5] Campbell, N., (2007) "Changes In Voice Quality with respect to Social Conditions", in Proc 16th ICPhS 2007.

[6] Hanson, H. M., (1995). "Glottal characteristics of female speakers". Ph.D. dissertation, Harvard University.

[7] Käre Sjölander, (2006). The Snack Sound Toolkit from http://www.speech.kth.se/snack/

[8] Pearson, K., (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space". Philosophical Magazine 2 (6): 559-572.

[9] R Development Core Team, (2004). "R: A language and environment for statistical computing". R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, http://www.R-project.org

[10] Chernoff H, Lehmann E.L.(1954). "The use of maximum likelihood estimates in χ^2 tests for goodness-of-fit". Annals of Mathematical Statistics, 25:579-586.

[11] Campbell, N., and Mokhtari, P., (2003). "Voice Quality is the 4th Prosodic Parameter". *Proc. 15th ICPhS* Barcelona, pp.203–206.

[12] Campbell, N., Kashioka, H., Ohara, R., (2005). "No laughing matter", pp.465-468 in Proc INTERSPEECH-2005.