

# ACOUSTIC FEATURES OF FOUR TYPES OF LAUGHTER IN NATURAL CONVERSATIONAL SPEECH

*Hiroki Tanaka<sup>a</sup> & Nick Campbell<sup>a,b</sup>*

<sup>a</sup>Nara Institute of Science and Technology, Japan;

<sup>b</sup>Trinity College Dublin, Ireland

hiroki-tan@is.naist.jp; nick@tcd.ie

## ABSTRACT

This paper presents the results of an analysis of the representative sounds of human laughter from a large corpus of naturally-occurring conversational speech. Two contrasting manners of laughter were categorized for the study: polite formal laughs and sincere mirthful laughs, and a formant analysis was performed on four phonetic classes of laugh therein. Laughing speech was also common in the corpus but is not addressed in this work. Statistical analysis of the acoustic features of each laugh was performed, and formant parameters were compared for each call type within a laughter bout. The paper details the formant characteristics and shows how polite laughter can be distinguished from sincere laughter on this basis using a trained statistical recogniser.

**Keywords:** spoken discourse, natural interaction, laughter, formant analysis, polite vs sincere laughs

## 1. INTRODUCTION

Laughter forms an essential component of human spoken interaction but has not been the subject of mainstream phonetic research until quite recently.

An analysis of a large corpus of spontaneous conversational speech recorded in highly natural situations revealed that laughter played an important role in all dialogues but the laughs only occasionally resulted from deliberate humour; polite laughter and nervous social laughter accounted for more than half the number of laugh bouts [2].

There is evidence that laughter is deeply rooted in human biology. “Based on acoustic form and likely phylogenetic history, laughter is argued to have evolved primarily as a vehicle of emotional conditioning. In this view, human laughter emerged because it helps foster and maintain positive, mutually beneficial relationships among individuals with genuine

liking for one another. It is predicted to as easily have the opposite role among those who do not” [5]. The most extensive study of the spectral properties of humorous laughter was performed by Bachorowski and colleagues [1] and was based on laugh bouts recorded from 97 young adults as they watched funny video clips. The research revealed a consistent lack of articulation effects in supralaryngeal filtering and reported formant related filtering effects which were found to be disproportionately important as acoustic correlates of laughter, sex, and individual identity.

At a satellite workshop of the 16th ICPhS in Saarbrücken, Szameitat, et al [6] reported a similar acoustic analysis of several kinds of laughter but employed professional actors to produce the laughs. They confirmed the Bachorowski et al finding that laughter syllables are predominantly formed with central vowels, and showed that compared to speech production, the first formant of laughter vowels is occasionally characterized by exceptionally high frequencies which may be the result of a wide jaw opening and/or pharyngeal changes in “pressed” voice, while the vowel elements during laughter showed a relatively stable individual pattern.

In a seminal study of the segmentation of laughs, Trouvain [8] suggests that we consider laughter as articulated speech, where at the low level there are sound segments that are either vowels or consonants. At the next higher level, there are syllables consisting of sound segments. The next higher level deals with larger units such as phrases which are made up of several syllables. Owren [5] recommends the term ‘bout’ for the longer sequence, and ‘call’ for the individual syllables; we will adopt that terminology in this paper.

Our principal goal in this study is to help autistic children distinguish types of laughter in terms of phonetic classes of laugh in conversational speech.

## 2. TYPES OF LAUGHTER

Table 1 shows the counts of laughs extracted from a 30-minute conversation from the ESP speech corpus [4]. We determined in all five categories of laugh (1: mirthful 2: politeness 3: embarrassment 4: derision, and 5: other) from independent labeling by eighteen people, and the kappa statistic using Cohen's method is 0.37 ( $p = 5.179e-10$  (signif)).

**Table 1:** Counts of laughs and non-laughs in a representative thirty-minute conversation between two Japanese males (JMA and JMB).

Type	Count	Prop	Cumulative Prop
not laughs	6999	none	none
2	472	54%	54%
1	244	28%	82%
3	107	12%	94%
5	49	5%	99%
4	4	1%	100%

The mirthful and polite laughs accounted for 80 percent of all laughs. Embarrassed laughs are difficult to distinguish acoustically from polite laughs, so we conflate them together here with 'polite'. The table shows counts of labels for both laughs and laughing speech, but we omit laughing speech because of its linguistic complexity. Therefore we categorize the data into two main types: hearty or mirthful laughs, labeled 'm' and polite laughs, labeled 'p' in the figures below.

## 3. ANALYSIS OF LAUGHTER

### 3.1. Data

The ESP-C corpus was recorded over several months, with paid volunteers coming once a week to talk informally and with no specific instruction as to content with specific partners in a different floor 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 socialize 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. The speech data were transferred to a computer, segmented into separate files, and transcribed manually. 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 phonetic Katakana orthography, wherever possible, alongside the use of the symbol [3].

### 3.2. Formant feature extraction

For this study, we selected one typical conversation and analysed the voice of one male speaker (JMA). Figure 1 shows a sample of his labeled speech, identifying different kinds of laugh. Four phonetic manners of laugh were identified for further annotation in this conversation, to distinguish nasal, ingressive, vocal and chuckle subtypes at the call level. Formant analysis was performed using the Tcl/Tk 'Snack' speech processing toolkit [7] which includes legacy code from David Talkin's Entropic Signal Processing System (ESPC) including *get\_f0* and *formant*.

**Figure 1:** Formant estimation in Wavesufer, showing labels for each call (nasal, voiced, chuckle, or ingressive). These raw estimates were post-processed as explained in section 3.2 of the text.

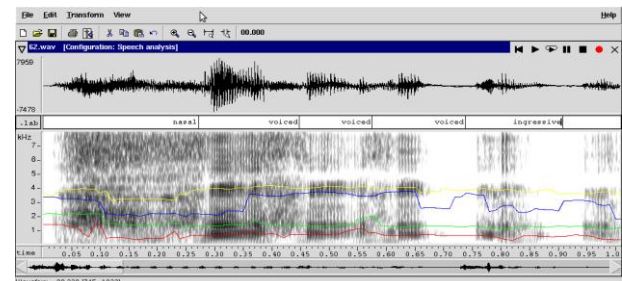


Figure 1 shows a sample of the formant estimation. It should be clear from the figure that the estimation is good for strongly voiced portions of the speech but can be almost random at other parts. We therefore devised a filter based on rms signal energy and thresholded the estimates to disregard all sections where speech power was lower than the overall mean determined for the signal.

### 3.3. Analysis results

The average fundamental frequency for each laugh was: 159 Hz (sd 30) for voiced laughs, 234 Hz (sd 64) for ingressives, and 197 Hz (sd 76) for

chuckles. Fundamental frequency estimation was not reliable for the nasal grunts.

The difference between the second and first formants (a useful measure of fronting) was voiced (v): 801 (sd 189), ingressive (i): 825 (sd 207), chuckle (c): 714 (sd 195), and nasal (n): 863 (sd 244). The three vowels, for comparison, were /a/ 714 (sd 204), /i/ 1440 (sd 352), /u/ 954 (sd 196).

**Table 2:** Counts and formant values (mean with sd in brackets) for each phonetic type described in the text.

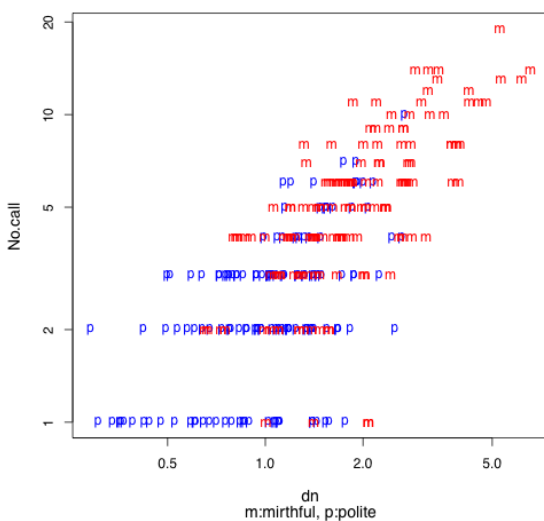
id	count	F1	F2	F3	F4
/a/	47	600(71)	1315(181)	2511(84)	3889(327)
/i/	38	443(98)	1883(280)	2760(278)	3821(260)
/u/	53	423(60)	1378(184)	2356(206)	3692(336)
v	88	643(138)	1445(165)	2575(215)	3559(208)
i	58	487(72)	1312(186)	2414(221)	3308(261)
c	69	500(204)	1215(239)	2136(378)	3022(435)
n	59	617(299)	1481(280)	2346(450)	3384(529)

**Table 3:** Confusion matrix (predictions) from HMM prediction trained on spectral features, testing unseen data.

	voiced	ingr.	chuckle	nasal	%c
voiced	39	0	2	0	95.1%
ingressive	0	10	7	0	58.8%
chuckle	1	1	24	3	82.8%
nasal	0	0	0	19	100%

Figure 2 shows how the polite laughs are distinguished from mirthful laughter in duration and number of calls. Polite laughter includes voiced laughs that sound similar to mirthful laughs, as well as nasal grunts ('hmmh') but rarely includes chuckles and never the ingressive laugh [3]. Mirthful laughter tends to be longer and to include more calls per bout.

**Figure 2:** Plot of log duration calls per bout.



The summary statistics obtained as a result of the formant analysis are plotted in Figure 3 and 4. Figure 3 shows estimates of the formant averages for all instances of three vowels (/a/, /i/, /u/ extracted from the conversational speech) in the various formant spaces, alongside equivalent plots for the four types of laughter. The figures confirm with findings from spontaneous conversational speech the findings of previous researchers using acted or humorous laughter that these laughs tend to cluster in the 'schwa' space, in spite of being perceptually more vocalic (i.e., often transcribed as 'haha', 'hihi' or 'hoho').

**Figure 3:** Plots of three vowels (/a/, /i/, /u/ extracted from the same conversational speech and four types of laugh (i: ingressive, n: nasal, v: voiced, c: chuckle) in the various formant spaces (left: f1,f2, middle: f1,f3, right: f3,f4).

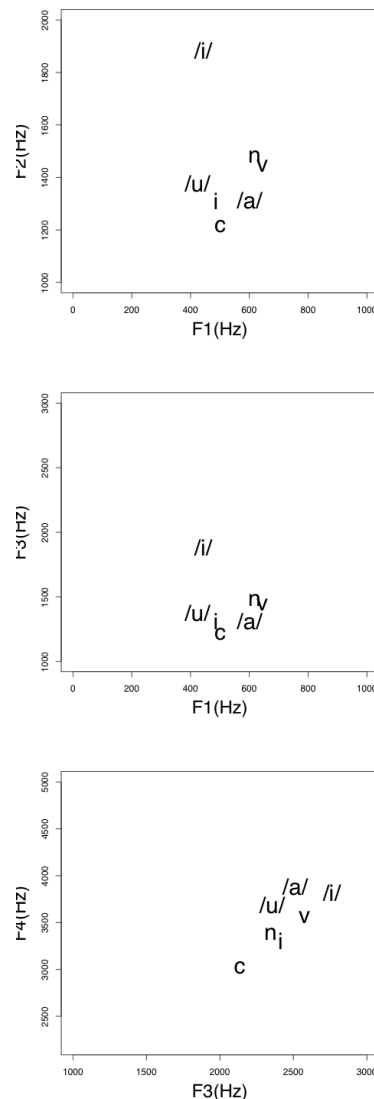
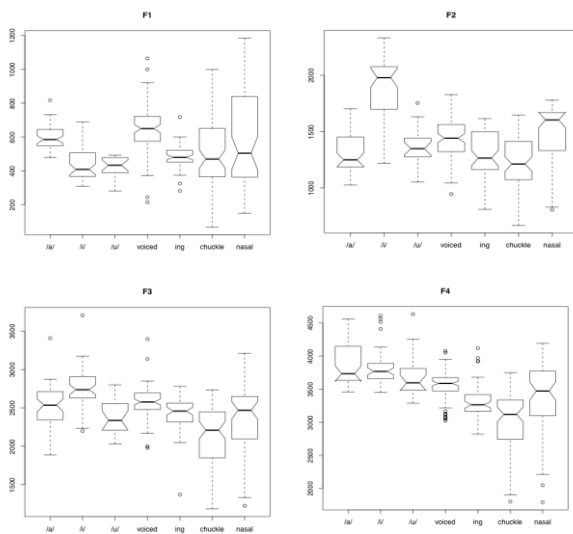
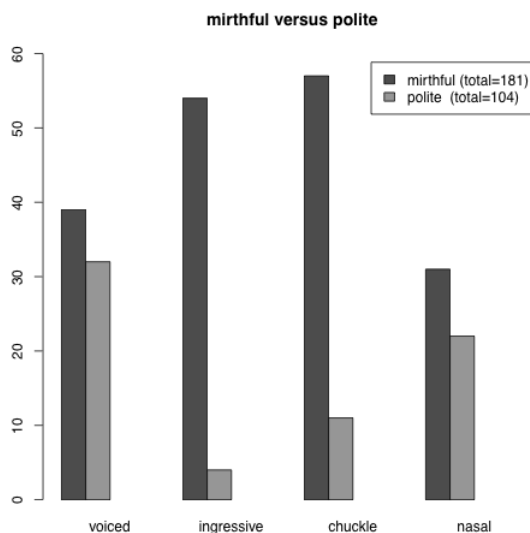


Figure 4 shows by means of boxplots the distribution within each formant space for the three vowels and four laugh call types. Figure 5 shows the distribution of each call type in the two manners of laugh. According to a Student's t-test between JMA and JMB to confirm generalizability of speakers, p-value is significant for voiced laughs' formant spaces (F1:  $p = 2.2e-16$ , F2:  $p = 3.0e-15$ , F3:  $p = 1.1e-15$ , F4:  $p = 2.9e-08$ ).

**Figure 4:** Showing distribution within each formant space for the 3 vowels and 4 laughs plotted in Fig 3.



**Figure 5:** Showing proportion of calls of each laugh type in polite and mirthful variants from a total of 26 bouts of laughter.



#### 4. STATISTICAL MODELLING

The discrimination of four types of phonetic classes of laugh enables us to classify manner of

laughter (i.e., ingressive laughs within a bout is likely to be sincere mirthful laughs). We applied a HMM statistical modeling based on a conventional phone recognition model to automated classification of each call type. To test the degree to which this spectral information can be used to distinguish between the different types of laugh, we trained an HMM using the following features: mfcc, rms power, and delta, power, with a subset of the laughs and tested on a different subset and achieved a prediction accuracy of 86.79%. (train: v 48, i 42, c 44, n 47), (test: v 41, i 17, c 29, n 19). Table 3 shows the confusion matrix and results from the prediction.

#### 5. CONCLUSION

Previous work reported formant frequency data for laughs using either acted samples or elicited humorous laughs. The present study reports findings from an analysis of natural spontaneous laughter in conversational speech which reinforce the findings of earlier studies. The present work extended the previous results to distinguish four types of phonetic laugh and two manners of social laughing. We were able to distinguish between these at levels much better than chance with a statistical detector trained on spectral features. This laughter detection is currently being integrated into a device to help autistic children distinguish between the two types of laughter.

#### 6. REFERENCES

- [1] Bachorowski, J.A., Smoski, M.J., Owren, M.J. 2001. The acoustic features of human laughter. *J.Acoust.Soc.Am.* 110, 1581-1597.
- [2] Campbell, N. 2007. Changes in voice quality due to social conditions. *Proc. 16th ICPhS Saarbrücken*, 2093-2096.
- [3] Campbell, N., Kashioka, H., Ohara, R. 2005 No laughing matter. *Proc Interspeech*, 465-478.
- [4] ESP (Expressive Speech Processing) corpus: <http://www.speech-data.jp>
- [5] Owren, M., 2007. Understanding acoustics and function in spontaneous human laughter. In Trouvain, J., Campbell, N. (eds.), *Interdisciplinary Workshop on The Phonetics of Laughter Saarbrücken*.
- [6] Szameitat, D.P., Darwin, C.J., Szameitat, A.J., Wildgruber, D., Alter, K. 2011. Formant characteristics of human laughter. *Journal of Voice* 25(1), 32-37. Tcl/Tk Snack Toolkit, [www.speech.kth.se/snack/](http://www.speech.kth.se/snack/).
- [7] Trouvain, J. 2003. Segmenting phonetic units in laughter. *Proc. 15th ICPhS Barcelona*, 2793-2796.