# ACOUSTIC ANALYSIS OF LAUGHTER

Corine Bickley[1] and Sheri Hunnicutt[2]*

[1]Research Laboratory of Electronics, Room 36-521, Massachusetts Institute of Technology, Cambridge, Ma. 02139 U.S.A.

[2]Department of Speech Communication and Music Acoustics, Box 70014, Royal Institute of Technology, S-10044, Stockholm, Sweden

*authors in alphabetical order

## ABSTRACT

Spontaneous laughter produced by two subjects was analyzed and compared to speech, measurements of both temporal and spectral characteristics being made. Speech and laughter were found to be quite similar in "syllable" duration and in the number of syllables per second. However, timing within syllables differed markedly. Fundamental frequency and rms amplitude of the laughter were also rather speech-like, although some extremes were observed. Formant structure of the laughter was also similar to speech. Bandpass filtering in the region of the third formant, however, showed the vocalic portions to include significant amounts of noise and breathiness, implying a more abducted vocal-fold configuration in the vocalic portions for laughter.

## INTRODUCTION

A need which has been revealed in the work to develop speech recognition algorithms is that of separating speech from non-speech. Of particular interest are those sounds which are acoustically close to speech due to being produced by the same physical system, the human vocal tract. Such sounds are coughs, sneezes, imitative sounds and other verbal effects or gestures as well as singing and laughter. The problem of separating these sounds from speech -- in order to recognize the speech -- is non-trivial. Both signals are created by the human vocal tract and thus may share both source and filter characteristics. In addition, these signals frequently co-occur. Laughed speech, for example, although we may hear little of it in our laboratory experiments, is quite common in everyday experience. Laughter co-occurring with speech was not addressed in this investigation, nor were the issues of speech variation due to emotion or smiling.

In order to begin an analysis of the separability of speech and laughter, it was decided to collect data containing both forms of vocal output from the same speaker. Before the collection was begun, however, recordings became available which were made for an unrelated speech error experiment. From these recordings we found two subjects which were suitable for our purposes. The spectrogram of a typical "laugh" by one of the speakers is shown in Figure 1. Our goal in this preliminary study was not to analyze all types of laughter or to examine a wide range of laughers, but rather to examine thoroughly a small set of laughs. It is recognized that there are many other kinds of laughter and that laughter is culture-related [1].

Because we were interested in comparing and contrasting our speech and non-speech materials on the basis of their speech-likeness, we chose measures typically used in speech analysis. Duration, frequency and amplitude characteristics were chosen to compare laughter with speech in general. Because laughter sounds like a sequence of breathy CV syllables, it also seemed interesting to compare laughter to reiterant /hV/ as in ha-ha-ha or heh-heh, and to compare particular characteristics of the breathiness of laughter with the breathiness in an /hV/ syllable. The analysis by Klatt and Klatt [3] of voice quality variations among female and male talkers provided the needed material for comparison. A further, applied comparison, will be made by synthesizing laughter in a formant (speech) synthesizer using the measurements from this study.

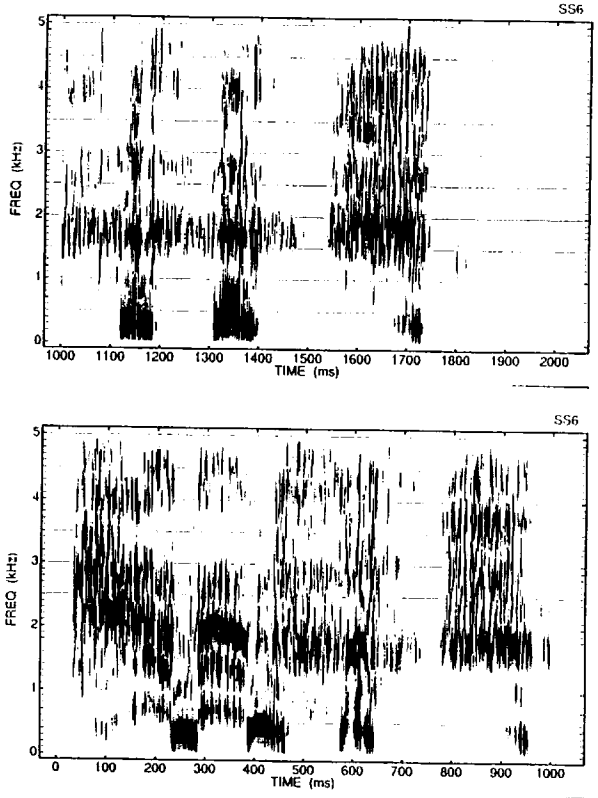


Figure 1. Spectrogram of a typical laugh. Note the alternating regions of unvoiced and voiced portions, the apparent aspiration in the unvoiced regions, and the breathy onset and offsets of the voiced regions.

## METHOD

For each laugh, several different acoustic measurements were made: durations of periodic and of noise-excited segments, fundamental frequency (F0) of the periodic segments, the difference in amplitude between the first and second harmonics of the periodic segments, the lowest three formants, the extra resonances (if any), the periodicity of the laugh in the region of the third formant, and the relative amplitudes of the syllables of each laugh. For the purposes of this study, a laugh syllable was defined as an unvoiced segment and a following voiced segment and a laugh defined as a sequence of laugh syllables.

### Subjects and materials

Laughs produced by two adult American English-speaking subjects, one female and one male, were used in this project. The laughs had been recorded during data collection for an unrelated study of speech errors. In each case, the laughter was the subject's spontaneous response to the process of recording the speech materials. The recordings were made in a sound-treated room, and were then digitized with a sampling rate of 10 kHz.

### Procedures

The following acoustic measurements of the temporal and spectral characteristics were made for ten laughs of one subject and five laughs of the other.

*Duration.* The waveform and spectrogram of each laugh were examined to identify the boundaries of the periodic, or voiced, segments and of the aperiodic, or unvoiced, segments. Low-frequency periodicity in the waveform was taken as evidence of voicing. The durations of each were calculated.

*Fundamental frequency.* The fundamental frequency of each voiced segment was derived by an algorithm [2] based on the spacing between the harmonics in the spectrum. In syllables in which the algorithm failed to identify a fundamental frequency, the inverse of the average period (over several cycles) was calculated.

*Difference in amplitude of lowest harmonics.* Narrow-band spectral sections of the laughs were computed without preemphasis at the beginning, middle, and end of each laugh syllable. A Hamming window of approximately 25.6 ms was used.

*Spectral peaks.* Formant tracks [2] and spectral sections were computed for each laugh. Spectral peaks were classified as resonances of the oral cavity or as resonances due to coupling to the trachea based on the continuity of the formant tracks and on comparison of the relative amplitudes of the peaks. The formant measurements were also needed for the synthesis component of this project, and for the filtering process which was part of the assessment of waveform periodicity.

*Waveform periodicity.* Each laugh was bandpass filtered (600 Hz) in the region of the frequency of the third formant. The resulting waveforms were examined visually and classified according to the amount of periodicity apparent on a four-point scale, following the procedure outlined by Klatt and Klatt.

*Waveform amplitude.* The rms amplitude at the midpoint of each laugh syllable was recorded, and the relative amplitude differences between syllables were calculated.

## RESULTS

The laughs produced by the two subjects can be described as sequences of alternating unvoiced and voiced segments (see Fig. 1). Timing, source characteristics, and vocal-tract configuration of the laughs were inferred from the acoustic measurements. The laughs produced by these subjects were occasionally mixed with speech as well as with snorts, choking sounds, and other non-speech-like behaviors; these sounds were not included in the results reported here.

### Timing

The average duration of a laugh syllable was found to be approximately 204 ms for one speaker and 224 ms for the other speaker; standard deviations were 36 and 32 ms, respectively. The average number of syllables per laugh was 6.7 for one speaker and 1.2 for the other. The periodic portion of each syllable was typically short in duration (the average value was 97 ms for one speaker and 68 ms for the other), and the unvoiced region was longer (averages of 167 ms and 156 ms for each of the two speakers, respectively). The ratios of unvoiced to voiced duration are thus 1.7 for one speaker and 2.3 for the other. These values are decidedly greater than a typical ratio of around .5 for spoken English [our own measurements].

### Source characteristics

In order to characterize the noise and periodic sources during laughter and to collect data for synthesis, several measurements including rms amplitude and fundamental frequency were made. For one subject, rms amplitude mid-syllable was found to vary between 40 and 70 dB with an average value of 53 dB. For the other subject, the range was 43 to 60 dB, with an average of 54 dB. The relative rms amplitudes within and between syllables were speech-like. These measurements formed the basis for specifying the amplitudes of the periodic and the noise-excited sources in synthesized laughter.

The range of F0 was observed to be 100 to 155 Hz for the male speaker with an average value of 138 Hz; for the female speaker the range was 161 to 476 Hz with an average value of 266 Hz. For the female subject, the values of F0 were higher than would be expected for her speech. The measurements of fundamental frequency for the male subject were speech-like in range and value. However, for neither subject did the F0 pattern throughout a laugh exhibit a typical pattern of declination often seen in speech. Figure 2 shows a plot of fundamental frequency corresponding to part of the laugh for the female speaker in Figure 1.

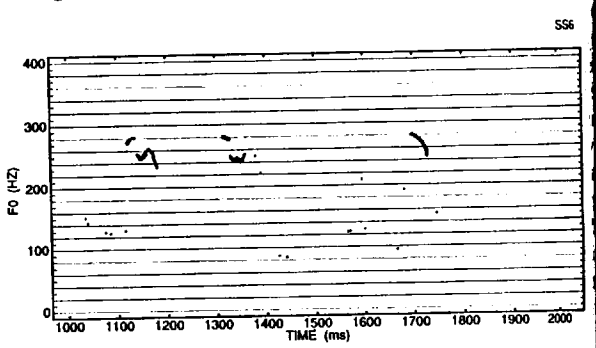


Figure 2. Plot of fundamental frequency of a typical laugh. N the speech-like variation. (The F0 contour indicated by the ha drawn line is based on measured values automatically generate 5-ms intervals.)

Following the procedure outlined by Klatt and Klatt we bandp filtered the vocalic portions of the waveforms in the region of third formant in order to determine the degree of random n present. We found that for most of the waveforms, aspiration n was commonly present in this region. Very few of the voc portions exhibited any significant periodicity (9%). Most of

portions (71%) were totally aperiodic. This amount of noise is not typical of speech. Figure 3 shows typical examples of waveforms judged as a) periodic b) somewhat periodic c) rather aperiodic and d) totally aperiodic.

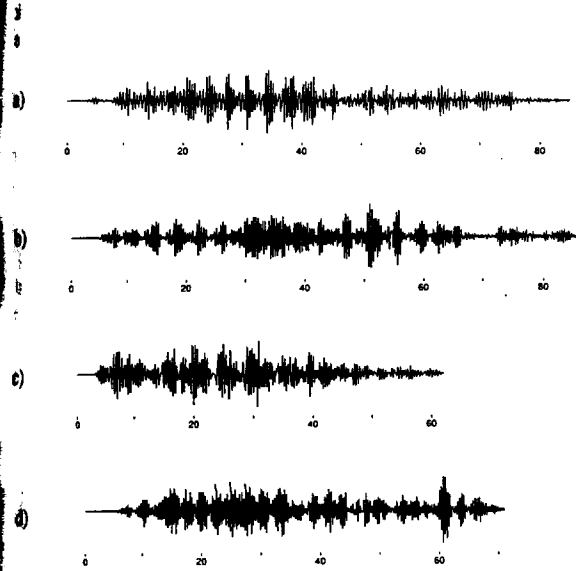


Figure 3. Samples of filtered waveforms of vocalic portions of laughs which were judged as a) periodic b) somewhat periodic c) rather aperiodic and d) totally aperiodic.

The shape of the glottal pulse as captured by the relative amplitude of the first harmonic was inferred at the edges and midpoint of each vocalic segment. Breathy onsets and offsets were apparent as evidenced by the greater relative amplitude of the first harmonic with respect to the amplitude of the second harmonic at the beginning and end of the vocalic segments. Breathiness was observed in 89% of syllable onsets, and in 86% of syllable offsets. These portions of breathy voicing imply an abducted vocal-fold configuration at the boundaries between the unvoiced and voiced segments of the laugh. Klatt and Klatt found similar breathiness in their reiterant /ha/ speech. In many of the laugh syllables (54%), the amplitude of the first harmonic was at least 2 dB greater than the amplitude of the second harmonic throughout the entire syllable; such prevalent breathiness is not common in speech. Figure 4 shows a sequence depicting the change in the amplitude of the lowest harmonics in the middle and at the offset of a laugh syllable in the spectrogram. Note the difference in the relative amplitudes of the first and second harmonics at mid-syllable compared to at syllable offset.

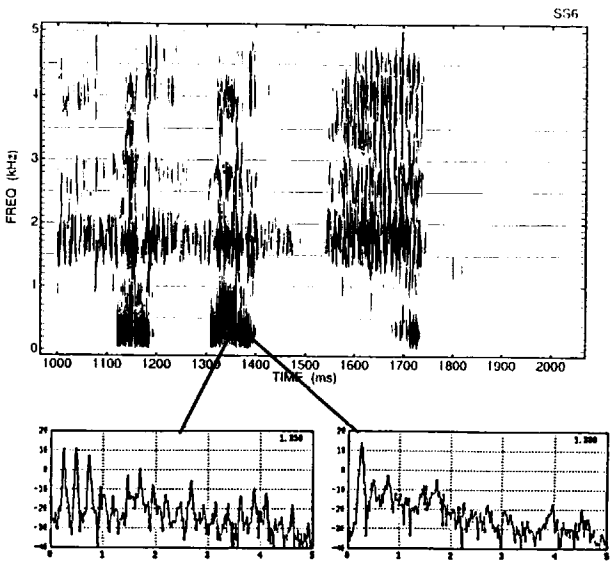


Figure 4. Spectrogram showing a typical breathy vowel onset and offset (from 1300 to 1400 ms). Below the spectrogram are two frames from the sequence of spectra at the vowel offset showing a radical change in the difference in the amplitudes of the first and second harmonics.

### Vocal-tract configuration

The frequencies of the lowest three formants were noted to be similar to those for vowels produced by the talkers. The spectrum on the left in Figure 5, showing the first vocalic portion in the spectrogram pictured above it, is typical. Here, the formant values are F1: 650, F2: 1700 and F3: 2200. The strong peak at 250 Hz is the first harmonic. Although the individual formant frequencies (F1, F2, F3) fall within the range of formants for each talker, the patterns of F1, F2 and F3 do not appear to correspond to a standard American English vowel. In several cases, values near F1 = 650 Hz, F2 = 1800 Hz and F3 = 2760 Hz were observed for the female speaker. One might hypothesize that she often used a particular vocal-tract configuration when laughing, i.e., this pattern of formant frequencies could signal a sort of "laugh vowel." For the female talker, extra resonances near 1000 Hz were found in many of the aspirated portions (see Figure 5, right-hand portion corresponding to the middle of the spectrogram above); for the male, the extra resonance was often seen around 950 Hz. These peaks could indicate coupling to the trachea [3].

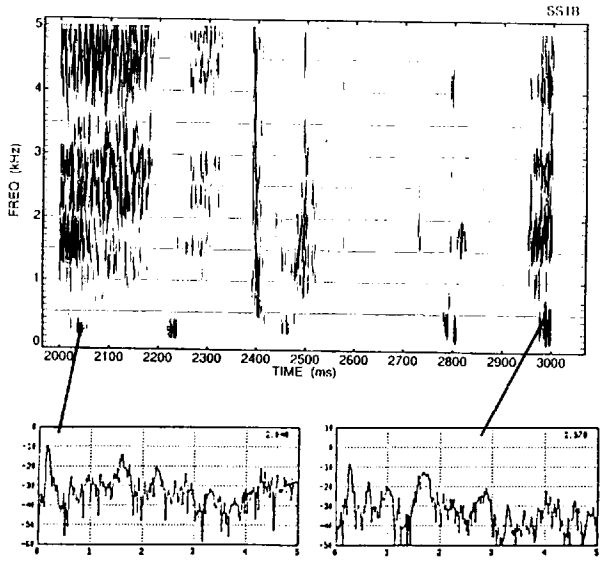


Figure 5. Sample spectra of vocalic portions of laugh to show formant structure. Note the extra resonance around 1000 Hz.

## DISCUSSION

Measurements of the temporal and spectral characteristics of the laughs indicated that in some ways laughter is speech-like: similarities include fundamental frequency range, formant frequency values, relative waveform amplitudes throughout a laugh, and breathy voice quality at the onsets and offsets of voiced portions of a laugh syllable. The average number of laugh syllables per second was calculated to be 4.7, which is similar to the syllable rates found for read sentences in languages such as Swedish, French and English (5.1, 6.1 and 4.7, respectively) reported by Fant, Kruckenberg, and Nord [4]. In other ways, laughs differ from speech; in particular, the durations of voiced portions are typically shorter in laughs than in speech, and the ratio of the durations of unvoiced to voiced segments is greater for laughter. Also, the glottal configuration for laughs appears throughout to be more abducted than for speech, as evidenced by the presence of noise in the region of the third formant, the enhanced amplitude of the first harmonic, and the frequent occurrence of tracheal resonances.

A possible method for separating laughter from speech, a "laugh detector," could be a scan for the ratio of unvoiced to voiced durations and for low-frequency voicing in periodic portions. Any waveform section with a large unvoiced-to-voiced ratio, particularly a series of these, or a long region which is heavily aspirated or breathy and cannot otherwise be accounted for as a sequence of lexical entries could be hypothesized to be laughter. This sequence could then either be ignored or could give a possible semantic clue to a joke or speech error.

## CONCLUSIONS

In many ways, the acoustic characteristics of laughter are speech-like (F0, formants, amplitude, voice quality). In particular, laughter is similar to reiterant /hV/ speech in terms of glottal characteristics, particularly at the boundaries between a vowel and /h/. However, the patterns of voicing in laughter and speech are quite different: laughs have significantly longer unvoiced than voiced portions and often have long regions of breathy voicing.


## ACKNOWLEDGEMENTS

Partial support for this work was received by the first author from NIH grant DC00075-29. The second author is grateful to Prof. Victor Zue for research support at M.I.T. during the scholastic year 1990-1991 when this work was in progress. Complementary work in 1992 has been done with the support of the Swedish Language Technology Program.


## REFERENCES


[1]Kori, S. (1987) "Perceptual dimensions of laughter and their acoustic correlates," Proceedings of the XIth International Congress of Phonetic Sciences, Tallinn, Se 67.4.1, p 255-258.

[2]Klatt, D.H. (1984): MIT Speechvax Users' Guide (unpublished).

[3]Klatt, D.H. and Klatt, L.C. (1990): "Analysis, synthesis, and perception of voice quality variations among female and male talkers.", J. Acoust. Soc. Am., 87:2, p 820-857.

[4]Fant, G., Kruckenberg, A. and Nord, L. (1991): "Durational characteristics of stress in Swedish, French and English.", J. Phon. 19, p 351-365.