

The speech laugh spectrum

Caroline Menezes, Yosuke Igarashi

The National Institute for Japanese Language
3591-2, Midori-cho, Tachikawa-shi, Tokyo Japan

Japan Society for the Promotion of Science

menezes@kokken.go.jp, yosuke.igarashi@kokken.go.jp

***Abstract.** This paper reports on the acoustic features of the laugh-speech continuum, specifically dealing with formant space, pitch range and voice quality. Results show that laugh vowels are generally reduced and there is a tendency for them to be produced more back and low in the vowel space when compared to speech. The large open quotient and shallow spectral tilt measurements for laughter is consistent with large sub-glottal pressure, more lateral compression of the glottis and greater glottal airflow. High sub-glottal pressure intrinsically raises fundamental frequency and could possibly explain the high F0 values we report for laugh and speech-laughs. Speech-laugh was found to be an intermediate between speech and laugh where some modulation of the opening phase occurs in order to speak, even with high sub-glottal pressure.*

1. General Information

Laughter is a special human vocalization chiefly because it induces a positive affective state on listeners. As more research is conducted on laughter more and more interesting facts and phenomena are revealed. Human babies begin to laugh at about 4 months, preceding early forms of babble, which begins at 6months-1 year of age (Sroufe & Waters, 1976). Even deaf and blind babies can laugh in the absence of auditory and visual feedback (Eibl-Eibesfeldt, 1989: referenced in Bachorowski et al., 2004). While laughter is universal there are many forms of human laughter. Bachorowski and colleagues investigating a large database of laughter sounds have differentiated three basic types namely, “song-like”, “snort-like” which contains nasal turbulence, and “grunt-like” which contains laryngeal and oral cavity frication. They also claim that only 30% of all analyzed laughs were voiced. According to Provine (1996) laughter in human communication does not interrupt the “speech stream”, but act somewhat as punctuation marks in written speech, i.e., proceeding or following phrases. However, it is difficult to imagine that laughter does not influence the acoustics of speech even if it does not disrupt the production of speech segments. Nwokah *et al.*, analyzing mother-child interaction find that about 50% of laughs in conversations co-occur with speech, what they refer to as “speech-laughs”, a simultaneous production of laugh and

articulation. Further they claim that “neither speech nor laughter is dominant when combined” and the complex vocal production of speech-laugh is a result of maintaining speech like fundamental frequency and laugh like amplitude and rhythm.

Ethologists have presented evidence for a direct evolutionary link between human laughter and chimpanzee vocalizations (van Hoof, undated web source, Provine, 1996). Owren & Bachorowski (2003) propose that the acoustics of both animal vocalization and laughter have direct influence on the affective state of the target animal (man in the case of laughter). In fact, Provine (1996) goes to the extent of stating that “the study of laughter provides a novel approach to the mechanisms and evolution of vocal production, perception and social behavior.” If human laughter shows developmental trends from primate vocalization is it possible that speech has developed from laughter? This is a big question? In this report we investigate the acoustic quality of speech and laughter and the intermediate state of speech-laugh as a small step in trying to understand the phonetic difference in speech and laughter.

2. Data

2.1 Data Collection

Eliciting spontaneous laughter is a difficult task, let alone eliciting a control data of spontaneous laughter and speech. For this reason samples were selected from a speech database collected by the second author for the investigation of intonation of Tochigi Japanese. Data comprised of 30 sentences with various syntactic structures and pragmatic contexts. The original goal was to study intonation so the sentences contained predominantly sonorants to avoid segmental perturbation on fundamental frequency contours. Since the recording was conducted using a portable recorder out in the field the two informants (whose data we report here) were both present during the session resulting in spontaneous dialogue and friendly laughter exchanges between them especially when they made mistakes performing the speech task. The informants were two female acquaintances from the same college.

The recordings were made using Marantz PMD 660 and saved onto a Compact Flash memory card at a 48 kHz sampling rate. Later they were down-sampled to 16 kHz and analyzed using the Wavesurfer program developed at the Royal Institute of Technology, Sweden.

The materials selected for this study were first classified into four categories, 1) prepared speech without laughter, 2) prepared speech with laughter, 3) spontaneous speech with laughter and 4) plain laughter. The first category is the control condition containing the test sentences read without laughter. The second category includes the test sentences produced with laughter. For this condition all sentences that contained laughter within or immediately preceding or following it were selected. Members of the first category always have corresponding members in the second category. The third category is defined as spontaneous speech. It includes utterances such as ‘*Oh, no!*’ or ‘*It’s terrible!*’ produced during the recording. Finally, the fourth category is defined as pure laughter produced without any speech.

2.2 Measurements

The goal of this study was to investigate differences in the acoustic features of laughter and speech by investigating formant frequencies, pitch range and voice source characteristics. All acoustic analyses were conducted using the Wavesurfer program. All categories except pure laughter were labeled into utterances, words, consonant and vowels. Laughter was identified as voiced laughter and unvoiced laughter. Voiced laughter was defined as laugh vocalizations that contained the fundamental frequency. For the purpose of this study only voiced laughter was analyzed.

The first three formants (F1, F2 & F3) were estimated from smooth spectral envelopes produced by linear predictive coding (LPC) analysis using Hamming window of 1024 sampling points, filter order of 16 and no pre-emphasis.

Spectral analyses were made using Fast Fourier Transform (FFT) using the same settings as the LPC. To compare the voice source characteristics of speech and laugh we looked at two glottal parameters namely, the glottal open quotient and spectral tilt. The extent of glottal opening during the glottal cycle affects primarily the lowest frequency components of speech signals (Stevens & Hanson, 1998). We measured this as the difference in amplitudes of the first two harmonics in the spectrum (H1-H2). On the other hand, the speed of vocal fold adduction in the glottal closing phase affects the middle and high frequency components and can be determined by analyzing the tilt in the spectrum (Stevens & Hanson, 1998). A gradual cessation of air flow during the closing cycle, therefore, results in steeper tilt of the spectrum in the high frequencies (for example in breathy phonation). Spectral tilt was measured as the difference between the amplitudes of the first harmonic and the peak harmonic in the third formant region (H1-A3).

Pitch was very high for laughter and sometimes difficult to obtain because of the irregular periods or strong glottal attack in the middle of the voiced segment. Therefore, we adopted a very simple process of selecting the frequency at the peak of the first harmonic (H1). For all syllables a point near the center of the acoustic vowel segment was selected for all acoustic measurements. This point coincided with relatively stable formants.

3. Results and Discussion

3.1 Formant Analysis

In order to analyze each vowel separately, the results were organized into three new categories namely, *speech*, *laugh* and *speech-laugh*. *Speech* and *laugh* correspond to the former categories of ‘prepared speech without laughter’ and ‘plain laughter’ (see 2.1 Data collection). The former categories of ‘prepared speech with laughter’ and ‘spontaneous speech with laughter’ were reclassified into *speech* or *speech-laugh*, depending on whether a given vowel is produced without laughter (*speech*) or with (noticeable influence of) laughter (*speech-laugh*). Nearly all the vowels in ‘spontaneous speech with laughter’ were categorized into *speech-laugh*. Thus, the new categories *speech*, *speech-laugh* and *laugh* form a continuum from speech to laughter.

In Figure 1, F1 and F2 values are plotted (pooled for both speakers) to show the vowel space map for *speech*, *speech-laugh* and *laugh*. Both axes are reversed so that the equivalence between acoustics and articulation are easier to grasp. Japanese is a five vowel system and contains no stress allophones¹. Also note that laughter, by definition, contains no linguistic information: there is no phonological specification of vowel categories. The vowel labels determined for *laugh* were made by auditory perception within the Japanese vowel system in order to identify the vowels of laughter.

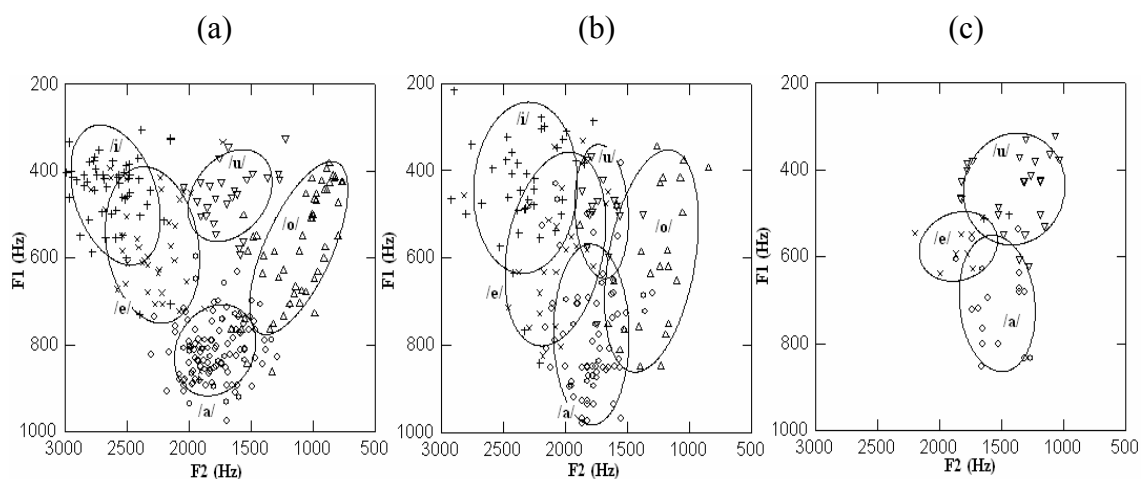


Figure 1. F1F2 plots for a) *speech*, b) *speech-laugh* and c) *laugh*.

In this figure we see how the vowel space is reduced as we go from *speech* to *laugh* (left to right). [See also Figure 2 with superimposed plots of the vowel space using average formant values]. However, most of the reduction occurs only in the F2 axis indicating that there is a trend for the tongue to retract as one laughs. The case of /o/ might appear contradictory but the slight increase in F2 could also be explained by lack of lip-rounding in speech-laughs. We prefer to use the term reduced instead of centralization for the trend between *speech* and *speech-laugh* because vowel identity was still maintained and there was no neutralization to schwa. Even in *laugh*, the

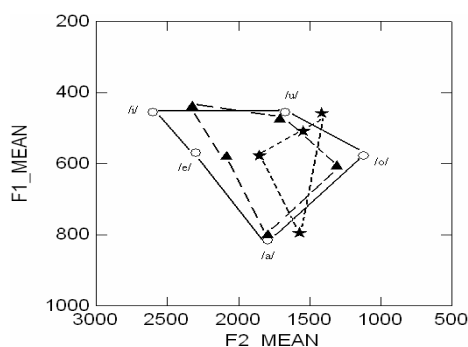


Figure 2. Vowel space for laugh (-o-), speech-laugh (-▲-) & laugh (-★-).

¹ Japanese /u/ usually lacks lip rounding and is centralized compared to IPA [u]. Acoustically, the centralization and weak rounding are manifested as relatively higher F2.

formant values exhibited a considerably wide distribution, so that laughter could be perceived in different vowel categories. Thus, the results for Japanese contradict the finding by Bachorowski and colleagues for American English where they report that laughter vowels are produced with an unarticulated vocal tract similar to that for schwa (Bachorowski *et al.*, 2001, 2004)².

We also find the increase of F1 for /e/, /o/ and /u/ in *speech-laugh* compared with *speech*, indicating a tendency to lower the jaw in speech-laugh. Notice in Figure 2, that the vowel space of both *speech* and *speech-laugh* is anchored at /a/. This means that laughter has little effect to reduce /a/ to mid vowel (schwa). Further, although the front mid vowel /e/ could easily be reduced to schwa by retracting the tongue back (reducing F2), some values of /e/ even approximate the F1 of /a/. Overall, the results along F1 scale also did not confirm the claim that laughter has an effect to reduce the vowels to schwa. Instead, the results suggest that laughter causes jaw lowering, resulting in an /a/-colored sound for all vowel categories.

So far, we saw putative effects of laughter on articulation, *i.e.*, lowering of second formant and raising of F1 related to the retraction of tongue and the opening of jaw. Interestingly, this trend can be noticed even in the vowel sequence used in a single laughter, *i.e.* along the temporal dimension. In our data, a bout of laughter can comprise of 1-3 laughs. In a typical laughter bout the initial (perceived) vowel could be /i, e, u/ but following vowels generally tended to be /a/ resulting in a /hihaha/ or /huhaha/ pattern (since there is no IPA symbol for laugh vowels we use the speech symbols that closely resemble the laugh vowels). That is, the tongue is retracted and the jaw is opened along the time scale as if the vowel were attracted to /a/.

3.2 Pitch Analysis

In table 1, we tabulate the descriptive statistics for fundamental frequency (F0) pooled for both speakers since both speakers showed similar trends. On average the fundamental frequency of laughter was very high when compared to speech. The result for laughter is not new as many have already reported very high F0 frequency for laughter when compared to speech (Bachorowski *et al.*, 2001, Bickley & Hunnicutt,

Category	N	Range	Minimum	Maximum	Mean	SD
Speech	285	366	148	514	242.65	62.88
Speech-laugh	178	372	166	538	273.75	76.32
Laugh	57	502	215	717	463.26	127.80

Table 1. Descriptive statistics for pitch pooled for both speakers.

1992, Mowrer *et al.*, 1987, Tartter, 1980). The trend for laugh F0 to be significantly higher than speech and speech-laugh are also reported by Nwokah *et al.*, (1999). Both peak F0 and F0 range exhibit a continuum between speech and laugh.

² The phonological system of Japanese does not include schwa.

3.3 Voice Quality Analysis

Figure 3 shows two measurements of voice quality separated by vocalization category and vowel type. On the left you see the values for open quotient and on the right the values for spectral tilt. The open quotient for *laugh* was significantly larger ($p < .001$) than *speech* and *speech-laugh* for all vowels. No significant difference was seen between *speech* and *speech-laugh*. Therefore, only for *laugh* the vocal folds were extremely spread during the open phase of the glottal cycle, but *speech* and *speech-laugh* were produced with the same glottal opening. On the other hand the spectral tilt plot on the right reveals that *speech* had the steepest spectral slope while for *laugh* the slope was the least. Further, there was a gradation in mean values across categories. Bonferroni test revealed significant differences between all three categories ($p < 0.001$). For a given open quotient, rise in the mid and high frequency amplitudes relative to the low frequency amplitudes occurs when the glottis experiences a quick closing phase (Klatt & Klatt, 1990, Stevens & Hanson, 1998).

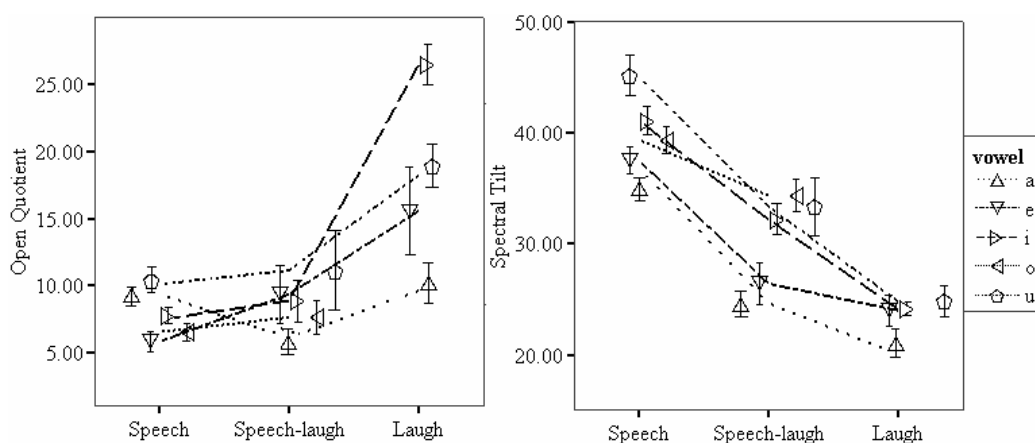


Figure 3. Open quotient and spectral tilt values separated by vocalization type and vowels.

Therefore, in *laugh* the glottal cycle begins with widely open vocal folds that are quickly and abruptly slapped together during the closing phase. It is known that high sub-glottal pressure causes greater lateral compression of the vocal folds and the quick airflow during the open phase increases the aerodynamic force that in turn causes the vocal folds to close rapidly. This high sub-glottal pressure could also be indicated in the high F0 values that we see for *laugh* and to some degree in *speech-laugh*.

The spectral envelope for *speech-laugh* depicts a combination between speech and *laugh*. Here the glottis is approximated typical to *speech* in the open phase, but the closing phase is more abrupt. It is possible that in *speech-laugh* the glottis opening is configured for air flow modulation just as for *speech* but the sub-glottal pressure is also high and this causes the rapid adduction of the vocal folds in the closing phase. During *speech* the subglottal pressure is maintained for a given level of loudness (Shadle, 1997).

Figure 4 gives an example of how the spectral slice looks for the vowel /a/ for a) *speech*, b) *speech-laugh* and c) *laugh*. [Note the example for *speech* is breathy phonation]. A dashed line is drawn at -50 dB to make it easy to compare the spectral tilt

across all three graphs. In here (Figure 4) another phenomenon special to laughter can be observed in the harmonic peaks. Harmonic peaks are not smooth as in *speech* but are jagged indicating that the amplitude of the noise spectrum is stronger than the periodic source spectrum. Normally, in speech the noise spectrum is below the periodic voice source in the frequency range below 5kHz (Stevens 1993, Shadle, 1997). This jagged edged harmonics indicate a strong turbulence noise source at the glottis during laughter, probably because the air flow during the open phase of the glottis is very high.

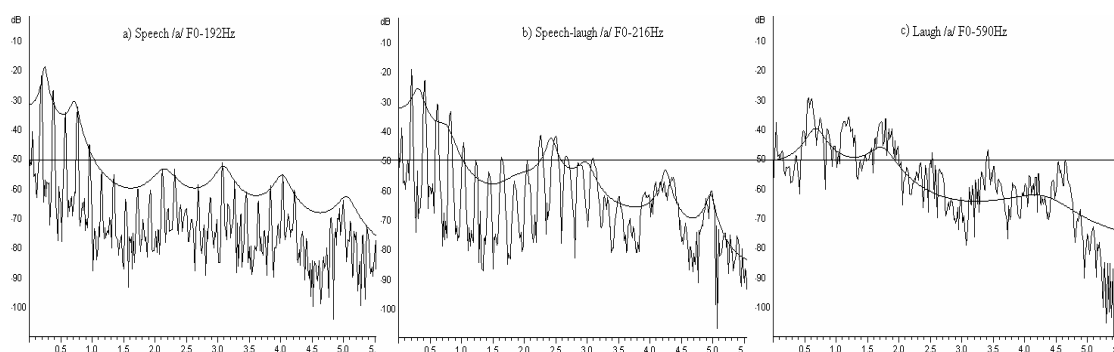


Figure 4. Spectral slice for speech, speech-laugh and laughter for the vowel /a/.

4. Conclusion

This study was conducted on a small data size as a preliminary study for future articulatory analysis. Indeed obtaining spontaneous articulatory data of emotive speech is no laughing matter at all. However, the results reported here indicate some interesting trends between speech and laughter. The vowel space for laughter is reduced and laugh vowels are more back and low when compared to speech, suggesting that laughter is produced with a lower jaw, a retracted tongue. However, articulatory analysis is needed to substantiate these findings especially for the case of /o/ and /e/. At this moment we do have articulatory data of laughter but analysis are still to be conducted. Results of the spectral analysis indicate that the glottal configuration for laughter is different from speech and speech laugh. The glottis is more abducted for laughter during the opening phase of the glottal cycle and the rate of airflow through the glottis is high. It is assumed that this is due to high sub-glottal pressure which could also explain the high fundamental frequency values we report for laughter. From these results we can see that laughter does not require the same fine motor control as speech both for vowel articulation and for controlling the expiratory breath stream.

References

- Bachorowski, J. A., and Owren, M.J., Laughing matters. *APA* 18 (9), 2004.
- Bachorowski, J. A., Smoski, M.J. & Owren, M.J., The acoustic features of human laughter. *Journal of the Acoustical Society of America* 111 (3), pp. 1582-1597, 2001.
- Bickley, C. & Hunnicutt, S., Acoustic analysis of laughter. *Proc. Intern. Conf. On Spoken Language Processing Banff* (2), pp. 927-930, 1992.

- Chiba, T., and Kajiyama, M., In *The vowel: Its nature and structure*. Tokyo-Kaiseikan Publishing Company, pages 115-131, 1941.
- Mowrer, D.E., Lapointe, L.L., Case, J., Analysis of five acoustic correlates of laughter. *Journal of Nonverbal Behavior* 11 (3), pp. 191-199, 1987.
- Nowkah, E.E., Hsu, H-C., Davies, P. & Fogel, A., The integration of laughter and speech in vocal communication: a dynamic systems perspective. *Journal of Speech, Language & Hearing Research* 42, pp. 880-894, 1999.
- Owren, M.J., and Bachorowski, J.A., Reconsidering the evolution of nonlinguistic communication: The case of laughter. *Journal of Nonverbal Behavior*, 27 (3), pages 183-200, 2003.
- Provine, R., Laughter. *American Scientist* 84-1(Jan-Feb, 1996) pages 38-47, 1996.
- Shadle, C., The aerodynamics of speech. In Hardcastle, W.J. and Laver, J., editors, *The handbook of phonetic sciences*. Blackwell Publishers, pages 33-64, 1997.
- Sroufe, L.A. & Waters, E., The ontogenesis of smiling and laughter: A perspective on the organization of development in infancy. *Psychological Review*, 83, pages 173-189, 1976.
- Stevens, K.N., and Hanson, H.M., Classification of glottal vibration from acoustic measurements. In Fujimura, O. and Hirano, M., editors, *Vocal Fold Physiology*, pages 147-70. Cambridge MA: Hilltop University Press, 1998.
- Stevens, K.N., Models for the production and acoustics of stop consonants. *Speech Communication* 13, pages 367-375, 1993.
- Tartter, V.C., Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception and Psychophysics* 27 (1), pp. 24-27, 1980.
- van Hoof, J., The evolution of laughter. Web source transcripts <http://www.uni-duesseldorf.de/WWW/MathNat/Ruch/PSY356-Handouts/The%20Evolution%20of%20Laughter.pdf#search='evolution%20of%20laughter'>