NO LAUGHING MATTER: AN INVESTIGATION INTO THE ACOUSTIC CUES MARKING THE USE OF LAUGHTER

Bogdan Ludusan, Petra Wagner

Phonetics and Phonology group, Faculty of Linguistics and Literary Studies, Bielefeld University
{bogdan.ludusan, petra.wagner}@uni-bielefeld.de

ABSTRACT

Laughter is a paralinguistic phenomenon widely used in human communication. Previous studies on laughter have mainly looked at its acoustic realization and its functions, leaving the context in which laughter occurs relatively under-studied. We intend to partially fill this gap by conducting an investigation into the acoustic cues that mark the use of laughter. We focus on the syllables preceding laughter and we explore several relevant spectral features. The results obtained on an American English corpus of conversational speech show anticipatory effects on the syllable immediately preceding laughter, including: a higher F1, a higher spectral center of gravity and a greater spectral standard deviation. We discuss these findings in terms of the individual variation present in laughter.

Keywords: laughter, context, acoustic cues, conversational speech.

1. INTRODUCTION

In recent years, we have seen an increased interest in the study of paralinguistic phenomena present in conversational speech (see [12] for a review). Among the paralinguistic aspects of speech, laughter plays an important role due, not only to its extensive presence in conversational data [4], but also to the various functions it plays [5].

Laughter has been studied from different perspectives over the years. For example, its organization, its production in context and its roles have been extensively investigated in sociolinguistics and pragmatics [6]. Speech technology instead, has mainly focused on the description of laughter from an acoustic point of view, both of the entire event (bout), as well as of its individual components (call). For an extended definition see [15].

The physical characterization of laughter has been extensively documented in the literature (e.g. [8, 10, 1, 13]), but only a very few studies have investigated the context in which laughter is produced [16, 3]. Vettin and Todt [16] analyzed a number of laughter contexts (preceding/following full phrases, following back-channels, silence or partner's laughter or speech) and have shown that more than 80% of the laughter bouts were used after the speaker's or partner's full phrase (in equal proportion). These findings were confirmed in [3], which looked at laughter context using higher organizational levels (topics) as the analysis unit. It found that laughter occurs more often at the end of a topic (and, implicitly, of a phrase), rather than at its beginning.

Although the effect of phrase context on the characteristics of laughter bouts has been investigated in [16], there are no studies, to our knowledge, that analyzed the effect of laughter on the acoustic properties of its preceding context nor studies which used analysis units shorter than the phrase. Here, we intend to partially fill this gap by conducting an investigation into the possible effects of laughter on the vowels immediately preceding it.

In order to decide which acoustic cues to include in our analysis, we reviewed the spectral changes associated with laughter. For example, Szameitat and colleagues [13] found increased F1 values for laughter vowels, compared to non-laughter vowels, which was explained by pharyngeal changes associated with "pressed" voice. Based on this result, along with other findings showing significant effects of "pressed" voice on other formant values [7], we included the values of the first four formants in our study. Furthermore, laughter is often associated with arousal, and the later was shown to have an effect on the spectral center of gravity [11]. We, thus, considered also a set of cues characterizing the general shape of the vowel spectrum, including: spectral center of gravity, spectral standard deviation, spectral skewness and spectral kurtosis. We will compare the vowels of words preceding laughter with the vowels of identical words (having the same phonetic pronunciation), not preceding laughter.

The paper is organized as follows: In Section 2 we introduce the dataset and the methods used in this study. Then, we illustrate the results obtained for the analyzed cues in Section 3. We conclude with a discussion of our findings, in relation to the current state of the field.

2. MATERIALS AND METHODS

The Buckeye corpus of American English [9] was used for the experiments conducted in this study. It consists of conversations between an interviewee and an interviewer about everyday topics. A total of 40 interviews (20 females, 20 males) were conducted and each interview lasted between 30 and 60 minutes. The interviews were orthographically transcribed, automatically aligned and manually corrected. The recordings contain spontaneous speech, exhibiting typical phenomena like hesitations, disfluencies or laughter, which were also annotated.

For the current study we considered all words which preceded laughter (marked as LAUGH in the corpus; we did not consider speech-laughs in our analysis), had no special label associated to it (words that have been only partially produced, contained noise, etc) and contained at least a vowel. Also, in order to be sure that any obtained effect is due to the laugh following the selected word tokens we filtered out the words preceded by another laughter event. For each of the selected word tokens that preceded laughter (further called *laughter tokens*) we searched for candidates corresponding to the same, non-laughter preceding, word. They had to have an identical phonetic transcription (as one word can have multiple pronunciations in the Buckeye corpus), be produced by the same speaker, not be preceded or followed by laughter and not be followed by a word containing special symbols. From the list of candidates, a randomly sampled word token was chosen to be the corresponding *non-laughter token*. If no corresponding non-laughter token was found for a given laughter token, the latter was removed from the analysis. The selected tokens were manually controlled and any error in segmentation or annotation was corrected.

From the total of 40 speakers, we considered for our study only those which had more than 15 laughter tokens. This resulted in 299 word tokens from 11 speakers (9 females). The vast majority of the tokens (90%) were monosyllabic words. For each of these tokens we extracted the following acoustic cues: the first four formants, the spectral center of gravity, the standard deviation of the spectrum, its skewness and its kurtosis. We employed in our study the average values of these cues, over the vowel of the last syllable of the token words. All the cues were extracted using Praat [2]. For the formants calculation we employed the burg method, by extracting the first five formants with a maximum formant value of 5500 Hz for female speakers and 5000 Hz for male speakers. The rest of the parameters used for extraction were set to their default values.

Table 1: Summary of the results obtained, for the eight cues analyzed in this study: F1, F2, F3, F4, spectral center of gravity (COG), spectral standard deviation (DEV), spectral skewness (SKE) and spectral kurtosis (KUR). For each cue we report its mean and standard deviation across the 11 speakers included in this study and the laughter/non-laughter t-test value ($^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$; $df = 10$ for each analysis).

Cue	Laughter		Non-laughter		t-test
	mean	stdev	mean	stdev	
F1	596.6	75.5	567.0	91.1	2.86 *
F2	1737	99.2	1738	109.2	-0.004
F3	2740	156.2	2749	172.0	-0.482
F4	3787	229.6	3846	263.0	-1.78
COG	634.0	127.8	558.5	90.7	2.64 *
DEV	504.1	109.1	441.9	63.5	2.51 *
SKE	3.559	0.897	4.113	0.905	-2.41 *
KUR	30.16	16.85	40.23	15.27	-2.64 *

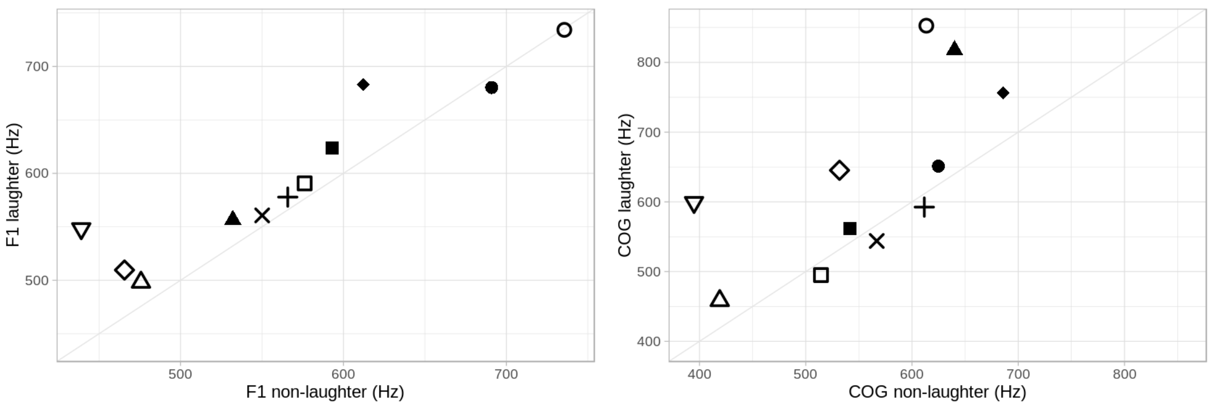
3. RESULTS

The results of the conducted analysis are presented in Table 1. For each of the investigated cues we illustrate their average value and standard deviation across speakers, for both laughter and non-laughter tokens, as well as the results of the paired, across speakers, t-test analysis comparing the two conditions.

It can be seen that F1 and the four spectral measures show a significant difference between the laughter and non-laughter conditions. F1, spectral center of gravity and spectral standard deviation are higher, while spectral skewness and kurtosis are lower. F4 decreases in laughter tokens, compared to the non-laughter condition, but the difference does not reach statistical significance.

Listening to a random sample of our data, we noticed a high degree of individual variation of laughter. Furthermore, it seems plausible that the type of vowel may have influenced the acoustic factors we are interested in. Therefore, we decided to verify the validity of the pairwise t-tests by running additional Linear Mixed Effects Models on all our dependent variables, with laughter and vowel type as fixed factors, and speaker as random factor (random slopes). These models confirmed a significant impact of laughter on the spectral characteristics of the vowel preceding laughter, with the vowel not preceding laughter having: lower F1 ($\beta = -28.4, SE = 7.17, t = -3.96$), higher F4 ($\beta = 49.88, SE = 19.81, t =$

Figure 1: The results obtained for F1 (left panel) and spectral center of gravity - COG (right panel) for tokens preceding laughter vs. tokens not preceding laughter. The speaker-averaged frequency for the non-laughter condition is displayed on the horizontal axis and the one for the laughter condition on the vertical axis. Each symbol represents a different speaker and the same symbols were used for the speakers in the two panels. The grey line represents the line for which the laughter and non-laughter conditions have equal values.



2.52), lower center of gravity ($\beta = -73.68, SE = 17.74, t = -4.15$), lower standard deviation ($\beta = -62.95, SE = 14.42, t = -4.37$), higher skewness ($\beta = 0.574, SE = 0.172, t = 3.33$) and a higher kurtosis ($\beta = 9.36, SE = 3.9, t = 2.4$).

Detailed results for F1 and spectral center of gravity are illustrated in Figure 1. Each point represents a different speaker included in our study and the values of the two cues in both laughter/non-laughter condition are presented as a scatter plot. We observe that the increase in the two cues is consistent across speakers: all, but two speakers, produce a higher F1 in laughter condition than in non-laughter condition – they are above the equal-value diagonal line, one (represented by an empty circle) showing no difference and another one (symbolized by a full circle) showing a small decrease. A similar trend is seen for the center of gravity: all, but three speakers (represented by empty square, x sign and cross), show an increase for the laughter tokens.

In order to investigate if the effects reported above extend also to the penultimate syllable of the word tokens, we considered all words which have at least two syllables and repeated the analysis for the vowel in the second-last syllable of the word. The results obtained showed similar trends to those observed for the vowel immediately preceding laughter, for most of the investigated cues. Still, the trends did not reach significance, probably due to the sparseness of our data (it included only seven speakers, with an average of four tokens per speaker).

4. DISCUSSION

We have presented here an investigation into the anticipatory effects of laughter on the preceding vowels. We use a per-speaker matched analysis, in which the last vowel of a word preceding laughter is compared to the last vowel of the same word, not found adjacent to laughter. We have looked at the values of the first four formants and at four different spectral measures: center of gravity, standard deviation, skewness and kurtosis.

Consistent with previous studies finding an increased F1 for laughter [13], we observed a similar increase in F1 for the vowel in the laughter token, compared to the one in the non-laughter condition. Furthermore, laughter token vowels seem to exhibit the characteristics of "pressed" voice, as described in [7], a higher F1, little change in F2 and F3, and a lower F4, although the latter was not found to be significant by the t-test analysis.

In terms of spectral cues, we found an increased center of gravity in the laughter condition. This effect is usually associated with arousal [11], which may well be the case for laughter tokens considered in this study. The remaining spectral features, although not generally associated with laughter, have revealed significant results. It seems that vowels followed by laughter deviate more from the center of gravity, but also have a more Gaussian shape around the center of gravity and a higher similarity between the two sides of the center of gravity (assumably due to the interaction between the increase in center of gravity and F2 and F3 getting closer together). It would be interesting to examine whether these characteristics apply also to actual laughter.

An intriguing observation can be made with respect to the results illustrated in Figure 1. While not all speakers exhibited an increased F1 or a higher center of gravity, all of them showed at least one effect. For example, the speakers not showing an increased F1, showed a higher center of gravity and the other way around. This is an encouraging result, as it means that the effect is consistent. At the same time, it points also to possible individual differences in the way speakers mark the use of laughter. It would be worthwhile to investigate whether the spectral effects we documented on the vowels preceding laughter may depend on the way speakers laugh or on other factors.

5. CONCLUSIONS AND FUTURE WORK

We investigated in this study the anticipatory effect of laughter on the vowel preceding it. We found a number of significant differences between the vowels of the same words preceding/not preceding laughter, for a range of spectral cues. These differences are consistent with spectral changes observed in the case of "pressed" speech and aroused speech. We see the current investigation expanding in three different directions. First, by performing a more detailed analysis of the observed effects, with respect to the different types of laughter the words precede (seeing how various types of laugh differ in their acoustic realization [14]). Next, we plan to extend our investigation to other, relevant, acoustic-prosodic phenomena, like phrasal prominence or intonation. And last, it would be worthwhile investigating the potential perceptual effect of the presence of the detected cues indicating laughter onset prior to its manifestation, i.e. whether listeners can anticipate laughter based on these cues, and perhaps join in promptly or show another adequate conversational reaction. Such knowledge could also be of potential use in dialogue systems, and add to a general understanding of the dynamics related to conversational laughter.

ACKNOWLEDGEMENTS

Bogdan Ludusan's work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 799022.

6. REFERENCES

[1] Bachorowski, J.-A., Smoski, M. J., Owren, M. J. 2001. The acoustic features of human laughter. *The Journal of the Acoustical Society of America* 110(3), 1581–1597.

[2] Boersma, P. 2002. Praat, a system for doing phonetics by computer. *Glot International* 5, 341–345.

[3] Bonin, F., Campbell, N., Vogel, C. 2012. Laughter and topic changes: Temporal distribution and information flow. *Proc. of the 3rd International Conference on Cognitive Infocommunications* 53–58.

[4] Campbell, N. 2007. Approaches to conversational speech rhythm: Speech activity in two-person telephone dialogues. *Proc. of XVIth International Congress of the Phonetic Sciences* 343–348.

[5] Dupont, S., Çakmak, H., Curran, W., Dutoit, T., Hofmann, J., McKeown, G., Pietquin, O., Platt, T., Ruch, W., Urbain, J. 2016. Laughter research: a review of the ILHAIRE project. In: *Toward Robotic Socially Believable Behaving Systems-Volume I.* Springer 147–181.

[6] Glenn, P. 2003. *Laughter in interaction* volume 18. Cambridge University Press.

[7] Laukkanen, A.-M., Sundberg, J., Björkner, E. 2004. Acoustic study of the "throaty" voice quality. *TMH-QPSR* 46, 13–24.

[8] Mowrer, D. E., LaPointe, L. L., Case, J. 1987. Analysis of five acoustic correlates of laughter. *Journal of Nonverbal Behavior* 11(3), 191–199.

[9] Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., Fosler-Lussier, E. 2007. Buckeye corpus of conversational speech (2nd release)[www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology. *Ohio State University (Distributor).*

[10] Provine, R. R., Yong, Y. L. 1991. Laughter: A stereotyped human vocalization. *Ethology* 89(2), 115–124.

[11] Sauter, D. A., Eisner, F., Calder, A. J., Scott, S. K. 2010. Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology* 63(11), 2251–2272.

[12] Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., Narayanan, S. 2013. Paralinguistics in speech and language–State-of-the-art and the challenge. *Computer Speech & Language* 27(1), 4–39.

[13] Szameitat, D., Darwin, C., Szameitat, A., Wildgruber, D., Sterr, A., Dietrich, S., Alter, K. 2007. Formant characteristics of human laughter. *Proc. of the Interdisciplinary Workshop on the Phonetics of Laughter* 9–13.

[14] Tanaka, H., Campbell, N. 2011. Acoustic features of four types of laughter in natural conversational speech. *Proc. of the XVIIth International Congress of Phonetic Sciences* 1958–1961.

[15] Trouvain, J. 2003. Segmenting phonetic units in laughter. *Proc. of the XVth International Congress of Phonetic Sciences* 2793–2796.

[16] Vettin, J., Todt, D. 2004. Laughter in conversation: Features of occurrence and acoustic structure. *Journal of Nonverbal Behavior* 28(2), 93–115.