Laughter Modulation: from Speech to Speech-Laugh

Jieun Oh, Ge Wang

CCRMA, Department of Music, Stanford University, Stanford, USA
jieun5@ccrma.stanford.edu, ge@ccrma.stanford.edu

Abstract

Laughing while speaking, also referred to as speech-laugh, occurs frequently in social conversations. In order to understand how laughter influences the acoustics of its co-occurring speech signal, we take a synthesis approach in designing an interactive system for artificial “laughter modulation”: users input an arbitrary speech signal, and the system processes the signal to yield acoustic patterns characteristic of a speech-laugh. Implemented using the ChucK audio programming language, our prototype allows for real-time manipulation of modulation parameters for rapid experimentation. This paper describes key components of our prototype, to be demonstrated at the 2013 Interspeech Show & Tell. This synthesis-based approach to speech-laugh may serve as a starting point for the future of affective machine-generated speech.

Index Terms: laughter synthesis, laughter modulation, speech-laugh, paralinguistics, ChucK, Show & Tell

1. Introduction

The expressive varieties of laughter and individual differences make it challenging to pinpoint the perceptually relevant ingredients of speech-laugh. Despite the prevalence of speech-laugh in everyday conversations, only a few studies have explored how laughter influences the acoustics of its co-occurring speech signal. Trouvain provides a concise overview on the phonetic aspects of speech-laugh, offering suggestions as to how laughter may “contaminate” speech [1]: speech-laugh are generally characterized by having a reinforced expiratory activity, occasionally with a tremor (vibrato) in voiced segments, with a possibility for increased pitch caused by smiling. Trouvain further describes how speech-laugh generally extend over two syllables, occurring in all positions of a phrase, and often followed by an isolated laugh. See [2, 3] for further descriptions on speech-laugh and their sequencing and functions in dialogues.

Beyond analysis of speech-laugh using speech corpora, attempts to synthesize laughter—whether isolated or synchronized with speech—provide additional insights on how we may naturally emulate laughter. For instance, Trouvain and Schröder experiment with adding laughter to synthetic speech to assess the effects on perceived social bonding and appropriateness of the laughter [4]. Sundaram and Narayanan’s two-level model for laughter captures the overall temporal behavior of a laughter episode based on a mass-spring system and uses standard linear-prediction-based analysis—synthesis to generate laughter calls [5]. In contrast, Lasarcyk and Trouvain explore imitating conversational laughter with articulatory synthesis and diphone synthesis [6]. Urbain and colleagues experiment with an audiovisual laughing machine by calculating similarities between laughter instances [7]. Finally, the authors have designed an expressive “instrument” for synthesizing isolated laughter based on a formant-filter model for singing synthesis [8].

2. Methodology

The system described in this paper, to be demonstrated at the 2013 Interspeech Show & Tell, is unique in its attempt to generate speech-laugh by modulating the input speech itself (as opposed to superimposing laughter on speech), and to do so in a way that is real-time user-controllable. Our system does not take into consideration any semantic or syntactic information of the speech input, and simply utilizes the acoustic signal to create a sensation of laughter in speech. This simplistic approach is motivated by our intuition that speech-laugh is characterized by a pattern of rhythmic exhalations and inhalations co-articulated with syllabic segments in speech.

In this section, we summarize how we segment input speech during the analysis stage and generate speech-laugh during the modulation stage; see Figure 1 for an overview.

2.1. Intensity envelope analysis for segmentation

The goal of the analysis stage is to segment the input speech at approximately its syllabic boundaries such that the segments can be used during the modulation stage to produce an effect of pulsed laughter exhalations. The rationale and background for performing such a segmentation is described in [9].

We assume that syllabic boundaries lie approximately at the local minima of smoothed intensity envelope of the signal. We implement a simple envelope follower using a one-pole filter applied to squared samples of the input signal. Ideally, the envelope follower should be just smooth enough to follow the syllabic-level intensity changes. Rather than algorithmically determining the correct smoothing level, we offer a graphical user interface for setting the pole position of the one-pole filter.

2.2. Control parameters for laughter modulation

At a high level, our implementation of “laughter modulation” integrates elements of amplitude, pitch, rhythmic, and tempo modulations, performed on the segmented units.
we allow users to specify the tempo of segment onsets.

exhibits rhythmicity around 4–6 pulses per second [10, 11, 12,

Pitch level, and release duration of segmented units.

ADSR envelope parameters: attack duration, decay rate, sustain
rate, along with PitShift for adjusting the pitch, are used
during the modulation stage to generate laughter effects.

Figure 2 shows the user interface for interacting with our
system. Changing the control parameters has immediate effects
on the resulting sound, encouraging users to discover effective
parameter values that yield natural-sounding speech-laughs.

4. Conclusion

We have presented our first attempt at interactively modu-
lating speech with laughter. Even though only a subset of
parameter settings yield convincing output, our prototype is
able to manipulate speech in a way that elicits a sense of
laughter-induced modulations. For examples of audio in-
put and output, see https://ccrma.stanford.edu/

5. References


[2] K. J. Kohler, “speech-smile, speech-laugh and their se-
quencing in dialogic interaction,” *Phonetica*, vol. 65, no. 1-2, pp.
1–18, 2008.

and speech laugh, based on observations of child-robot interaction,”

thetic speech,” in *Affective Dialogue Systems*. Springer, 2004,
pages 229–232.

human-like laughter,” *The Journal of the Acoustical Society of

with an articulatory speech synthesizer,” *Proceedings of the Inter-
disciplinary Workshop on the Phonetics of Laughter, 2007.*

C. Pelachaud, B. Picart, J. Tilmanne, and J. Wagner, “Avlaughter-
cycle: An audiovisual laughing machine,” in *Multimodal Inter-

cedings of the International Conference on New Interfaces for

level units in laughter,” in *Proc. Interspeech 2013, ISCA*

tures of human laughter,” *The Journal of the Acoustical Society of

between laughter and speech,” *Speech Communication*, vol. 49,

930.


strongly-timed and on-the-fly environment/mentality,” Ph.D. disserta-
tion, Princeton University, 2008.

New interfaces for ChucK development and performance,” in *Pro-
cedings of the International Computer Music Conference*, 2006,
pp. 63–66.

Figure 2: User interface for interactive laughter-modulation

**Amplitude:** The intensity contour of syllabic-level units have
been shown to be an important cue for automatically detecting
laughter in speech [9], and we therefore allow specification of
ADSR envelope parameters: attack duration, decay rate, sustain
level, and release duration of segmented units.

**Pitch:** Speech-laugh may result in increased pitch caused by
smiling [4], and the fundamental frequency of laughter may
have greater variations and extend higher than that of speech
[10, 11]. Motivated by this, we allow users to specify the level
of pitch modulation, which determines the upper bound on how
much the pitch may rise. As described in [8], we implement a
slight bending down of pitch in the latter portion of a segment
to induce characteristic qualities of laughter.

**Rhythm:** We hypothesize that the rhythm of syllabic unit onsets
would become more uniformly spaced in a speech-laugh con-
text, during which speech is synchronized with the exhalation
pulses. Thus we incorporate a parameter for rhythmic modu-
lation, which specifies the degree to which segmented speech
units are snapped to isochronous rhythm.

**Tempo:** Finally, motivated by empirical findings that laughter
exhibits rhythmicity around 4–6 pulses per second [10, 11, 12, 13],
we allow users to specify the tempo of segment onsets.

3. System

Our system is implemented using the ChucK programming lan-
guage [14] with MAUI Elements for the user interface [15].
ChucK provides a platform for precise audio synthesis/ an-
alysish and rapid experimentation – which is quite appropriate for
our purpose of synthesis-based explorations on speech-laugh.

The analysis- and modulation-patches used in the system
are comprised of built-in ChucK unit generators (UGens). LiSa’s,
a live sampling utility, is used to record input speech and
playback processed audio. The microphone input (adc)
connected to Gain multipliers followed by a OnePole fil-
ter constitutes a simple envelope follower used in our analysis
stage. LiSa’s built-in controls for gain envelopes and buffer
rate, along with PitShift for adjusting the pitch, are used
during the modulation stage to generate laughter effects.

Figure 2 shows the user interface for interacting with our
system. Changing the control parameters has immediate effects
on the resulting sound, encouraging users to discover effective
parameter values that yield natural-sounding speech-laughs.