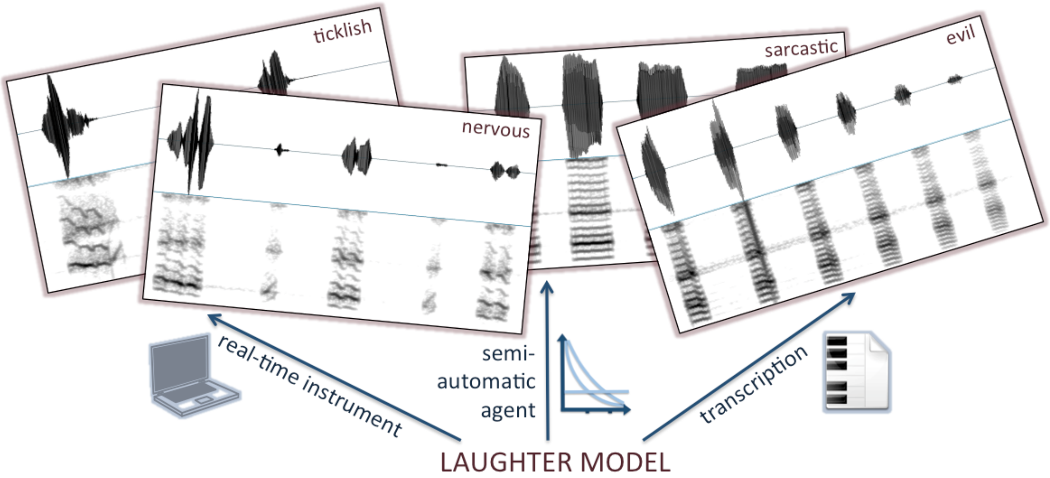
LOLOL: Laugh Out Loud On Laptop

Jieun Oh and Ge Wang
Center for Computer Research in Music and Acoustics (CCRMA)
Stanford University
660 Lomita Drive, Stanford, CA, USA
{jieun5, ge}@ccrma.stanford.edu



ABSTRACT
Significant progress in the domains of speech- and singing-synthesis has enhanced communicative potential of machines. To make computers more vocally expressive, however, we need a deeper understanding of how nonlinguistic social signals are patterned and perceived. In this paper, we focus on laughter expressions: how a phrase of vocalized notes that we call "laughter" may be modeled and performed to implicate nuanced meaning imbued in the acoustic signal. In designing our model, we emphasize (1) using high-level descriptors as control parameters, (2) enabling real-time performable laughter, and (3) prioritizing expressiveness over realism. We present an interactive system implemented in ChucK that allows users to systematically play with the musical ingredients of laughter. A crowdsourced study on the perception of synthesized laughter showed that our model is capable of generating a range of laughter types, suggesting an exciting potential for expressive laughter synthesis.

Keywords
laughter, vocalization, synthesis model, real-time controller, interface for musical expression

1. MOTIVATION
Over the past decades, computers have been made to speak [11, 21] and sing [5, 19]. Perhaps the next challenge for them, in their pursuit of vocalizing as humans do, is to perform nonlinguistic expressions that naturally reflect emotions. Various social signals, from gasps and sighs to yawning and sneezing, serve as important cues to our affective states. But among these, laughter stands out as being unusually musical in its expressiveness and variety.

Although funny jokes and comical situations do make us laugh, laughter triggered by ludicrousness comprises only a small proportion of laughter we experience everyday [23]. In fact, the social functions of laughter cover diverse scenarios, including displaying affiliation, aggression, fear, anxiety, joy, and even sadness [22]. Laughter triggered by different scenarios is characterized by distinctiveness in auditory features that implicate certain state and attitude of the laughing person. That is, we are able to label a sound of laughter as, say, "stealthily evil" or "out-of-control ticklish". This phenomenon makes us wonder how musical patterns of laughter lead to expression.

Given such expressive variety of laughter, our ultimate goal—the larger context behind the research described in this paper—is to understand how social and emotional meaning arises from a phrase of appropriately shaped vocalized notes that we call "laughter". Our methodology is through interactive musical synthesis and performance; the model described in this paper is the first (to the best of our knowledge) to realize real-time performable laughter.

2. APPROACH
In his research on vocal modeling, Perry Cook underscores a difference in priorities between synthesis of singing and speech: intelligibility is the primary goal in speech and speech synthesis, while quality is the main goal in singing and singing synthesis (often compromising intelligibility) [5]. Now, the prospect of synthesizing laughter forces us to re-identify our priorities. A laughing agent probably does not need to make distinctions in subtle phonemic differences that are necessary for speech recognition, nor does it need to strive for aesthetic quality in sustained tone production.

NIME'13, May 27 – 30, 2013, KAIST, Daejeon, Korea.

The agent should, however, preserve perceptually relevant characteristics of laughter that trigger certain social, affective responses from listeners. Thus we identify our priority in laughter synthesis as *expressiveness*, operationalized as the potential to convey a desired social and emotional meaning through the auditory signal. In this project we assume the motto of "expressiveness trumps realism."

3. RELATED WORKS

3.1 Expressive Speech and Singing Interfaces

There exist a great number of studies that contributed to the synthesis of vocalized sounds, from speech to singing (and, to lesser degree, laughter). Here we highlight just a few that particularly emphasize controllable expressivity.

3.1.1 Voder

The voder is a complex machine for vocal synthesis modeled after the human vocal tract. Developed by Homer Dudley and demonstrated at the 1939 World's Fair, the voder allowed trained technicians to manually "perform" speech utterances [10]. A wrist bar, a foot pedal, and ten finger keys controlled the buzz/hiss selection, pitch, and gains of band-pass filters, respectively [13]. A few operators who eventually attained "virtuosity" could imbue sounds with inflections in such a way that made synthesized speech humanly expressive. For example, the voder could respond "*She* saw me" to the question "Who saw you?", and respond "She *saw* me" to the question "Did she see you or hear you?". The voder challenges us to design an interface for laughter that has comparable expressive potential but is easier to control. This would require understanding the invariant features of laughter and determining controllable parameters that result in perceptually distinct laughter sounds.

3.1.2 Pattern Playback

In order to better understand laughter, we may learn from explorations carried out by Franklin Cooper and colleagues in early 1950s, on the perception of synthetic speech sounds [9, 18]. The experiments using pattern playback (a machine that synthesized speech by converting pictures of the acoustic patterns of speech to sound [8]) led to discoveries in important acoustic cues for speech perception. One study identified sixteen two-formant patterns that are closest to the IPA cardinal vowels [9], and another study analyzed the shapes of formants in consonant-vowel transitions [18].

3.1.3 SPASM

In the domain of music, SPASM (Singing Physical Articulatory Synthesis Model) by Perry Cook realizes machine singing [5, 6]. The graphical user interface in SPASM allows users to modify various control parameters, including time domain waveform and spectral content of the glottis, the radius of each vocal tract section, the velum opening size, noise spectrum and location for tract turbulence, and much more. Because the model takes as inputs physical parameters, users can have an intuition for how to improve if the synthesis does not sound correct [6]. In this way, providing input parameters that are consistent with the user's mental model of the sound's construction is crucial, and we try to follow this principle in designing our synthesis interface.

3.1.4 SqueezeVox

In order to naturally and quickly control the singing model and thereby *perform* singing, Cook and Leider developed squeezeVox, an instrument based on an accordion paradigm [7]. The squeezeVox provides intuitive controls for pitch, breathing, and articulation through a keyboard with linear sensing strip; bellow motion with sensors that monitor air pressure; and additional buttons, sliders, accelerometer, and trackpad. Cook explains how the challenge in designing a universal controller for vocal synthesis stems from a myriad of control parameters that must be specified, regardless of what synthesis model we choose[1] [7].

3.2 Understanding Laughter

Even though a majority of the studies on human laughter take a theoretical approach (exploring in great depth *why* and *when* humans laugh), some researchers have made valuable empirical findings on *what* laughter sounds like. In this section, we highlight studies that look at the acoustic correlates of laughter and summarize approaches that have been applied to synthesizing laughter.

3.2.1 Acoustic Correlates of Laughter

As described earlier, the social functions of laughter cover diverse scenarios, from joy to fear and sadness [22]. Consistent with this observation and our assumption that much information about the laughter context is encoded in the acoustic signal, a few studies have tried to link perceived emotions in laughter to their acoustics.

Using acted laughter and subjective responses, Kori determined two primary perceptual dimensions of laughter to be <pleasant-unpleasant> and <superior-inferor> [15]. Kori found that "duration of the strong expiratory noise at the beginnig of laughter" was strongly correlated with the pleasantness dimension, while "interval between vowels", "F0 maximum or mean value", and "the rate of overall vowel amplitude diminishment" were strongly correlated with the superiority dimension. Put in musical terms, Kori's finding suggests that the duration of an initial laugh-note, interonset intervals, pitch contour, and dynamic changes across a laugh-phrase are important features for classifying laughter.

Szameitat and colleagues conducted a related study, investigating the acoustical correlates of laughter expressing four emotions: joy, tickling, taunting, and schadenfreude [31]. Using laughter produced by professional actors, the researchers found that the different emotions can be classified accurately (84%) from acoustical parameters, with prosodic parameters providing stronger discriminative power over vowel quality. For instance, tickling laughter was rapid and high-pitched; joyful laughter was rich in low-frequency energy and had the longest time between bouts; and taunting laughter had the lowest f0. Again, these findings suggest that pitch and rhythm are important features for laughter classification. We apply these insights to defining perceptually relevant control parameters for our model.

3.2.2 Synthesizing Human-like Laughter

There are only a few studies that have tried to synthesize human-like laughter. In 2006, Sundaram and Narayanan designed a two-level laughter model for automatic synthesis [29]. One level captures the overall temporal behavior of a laugh-phrase using the simple harmonic motion of a mass-spring system. A second level employs a standard linear prediction based analysis-synthesis method to synthesize laugh-note. This architecture allows users to define the control parameters for the harmonic motion (first level), as well as for the overall variation of pitch, amplitude envelope, and LP coefficients (second level). Implemented in Matlab, the system receives user-defined inputs at runtime. Subjective rating showed that naturalness of synthesized laughter is significantly below that of real human expressions; synthesis of realistic laughter remains a challenging problem.

[1] Synthesis models include formant[14, 24], LPC[1, 16], FM[3, 4], FOF[25], sinusoids+noise[27], or physical[5].

Parameter	Description	Possible Values
Rhythm	note onset timings, note durations, IOI (in sec. or ms.)	(free)
Pitch (f0)	fundamental frequency of voiced components in Hz (includes subtle random variations)	(free)
Pitch Bending	bending down of pitch upon release of exhale notes	on | off
Voicedness	the extent to which a laugh-note is harmonic vs. noisy	[0.0, 1.0], noisy to harmonic
Vowel Space	first and second formant of voiced component (coordinates in 2D vowel-space)	x [-1.0, 1.0] y [-1.0, 1.0]
Inhale vs. Exhale	whether the laugh-note is an exhalation (affects pitch contour and release duration)	true | false
Glottal Waveform	shape of the waveform: lower values (<1) are flowy; higher values (>1) are pressed	approximately 0.5 to 2.0
Decay Rate	how quickly the laugher returns to equilibrium (for semi-automatic mode)	tends to be a small +number
Threshold	the amount of intensity required to vocalize a laugh-note (for semi-automatic mode)	(0.0, 1.0)

Figure 1: Summary of control parameters in our synthesis model

In contrast to the laughter-specific modeling by Sundaram and Narayanan, Lasarcyk and Trouvain (2007) tested whether existing speech synthesis techniques can be more directly applied to laughter [17]. Laughter was modeled using articulatory synthesis and diphone synthesis, though the latter was shown to be limited because breathing or certain laugh syllables are not available in the predefined phones used for speech. Articulatory synthesis allowed for modeling the articulation process directly, although researchers also experienced some technical limitations (such as inadequate upper limit for pulmonic pressure) and limitations arising from our lack of understanding of laughter physiology.

4. METHODOLOGY

As illustrated by our survey of synthesis techniques, there is currently no obvious approach to modeling laughter. Given our priority in achieving expressiveness, our methodology will contrast prior works in following ways:

(1) Use **higher-level descriptors** as control parameters (often borrowing musical terminology), hiding from the user's awareness lower-level parameters whose effects on the resulting sound is not immediately perceivable

(2) Be **real-time controllable**, allowing users to manually perform laughter by triggering and adjusting synthesis parameters on-the-fly

(3) Prioritize **preserving musically salient features** of laughter over attaining realism

4.1 Control Parameters

Figure 1 summarizes high-level descriptors that define our model. These parameters have been chosen with hopes that they are easy to understand for users who do not have expert knowledge on the vocal tract or voice synthesis, and yet can represent musically salient characteristics of laughter.

4.1.1 Rhythm

The rhythm of laughter – including note onset timings, note durations, and interonset intervals – is explicitly definable in terms of time duration. A study by Bachorowski and colleagues [2] provides insightful analysis on the patterns of call durations (laugh-note durations) and intercall intervals (interonset intervals) that can be applied to our model.

4.1.2 Pitch (f0)

The pitch patterns of laugh-notes shape the melodic contour of a laugh-phrase, and this can be defined in terms of the fundamental frequency of laugh-notes.

4.1.3 Pitch Bending

Additionally, the pitch of a laugh-note should change throughout the attack, sustain, and release stages to reflect natural variations in human laughter. We incorporate continuous subtle variations (randomness) in the pitch of a laugh-note, and further enable bending down the pitch during the note-release stage, as we have found this to be an important feature that gives a "laugh" quality to an otherwise flat-sounding note (see Figure 2).

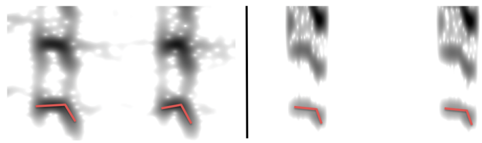


Figure 2: Spectrograms of real laughter (left) and our synthesized laughter (right) with pitch bending

4.1.4 Voicedness

The extent to which laughter sounds noisy versus harmonic is encapsulated in the voicedness parameter, ranging from 1 (voiced) to 0 (noisy). This parameter has been motivated by our observation that the stochastic components greatly influence our percept of the nature of laughter. Figure 3 shows how the spectral component of laugh-notes become smudged as we decrease the voicedness parameter.

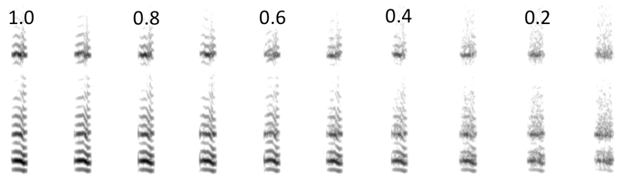


Figure 3: Effects of changing the voicedness parameter on harmonicity of synthesized laughter

4.1.5 Vowel Space

There are conflicting views on the importance of formants in a laughter context. Some argue that formants do not provide significant discriminative power for classification [31], while others suggest that a higher first formant may be indicative of a more extreme articulation during laughter production [32]. Perhaps this disparity can be partially resolved by specifying whether formants are summarized in absolute terms or tracked in their relative movements across time.

Based on our experience, continuously changing formant values—within a laugh-note as well as across laugh-phrase—greatly affects expression, contributing to a more human-like quality. Thus, our model provides an abstraction for specifying the spectral peaks in terms of a two-dimensional representation of the vowel-space. The x-coordinate roughly translates to the second formant, from back vowels to front vowels; the y-coordinate roughly translates to the first formant, from low vowels to high vowels. See Figure 4 for how changing the coordinates affects spectral peaks in laugh-notes, contributing to vowel coloring.

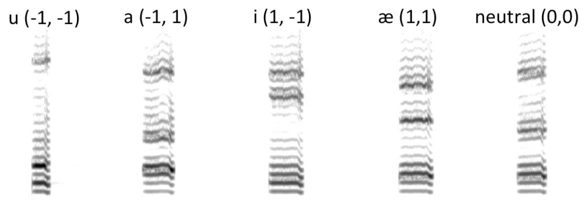


Figure 4: Effects of vowel space coordinates on synthesized formants

4.1.6 *Inhale vs. Exhale*

One significant way in which laughter differs from speech or singing is in the expressive power of the inhalation gesture that often marks the end of a laugh-phrase. For instance, an abrupt inhalation may imply nervousness, and a voiced inhalation may imply certain lack of control. Our model includes a parameter for specifying inhalation, and inhale notes have different default pitch contour and release duration than exhalation notes, as shown in Figure 5.

Figure 5: Comparison of synthesized exhalation (left) and inhalation (right)

4.1.7 *Glottal Waveform*

A parameter for the glottal waveform shape offers users control over the voice timbre that result from harshness of glottal closure. This parameter has been inspired by how the different types of phonation methods humans use, from "breathy" and "flowy" to "neutral" and "pressed" voices, result from the changes in the shape of the glottal waveform [26, 30]. See Figure 6 for the effects of changing waveform shape on the higher frequency components of laugh-notes.

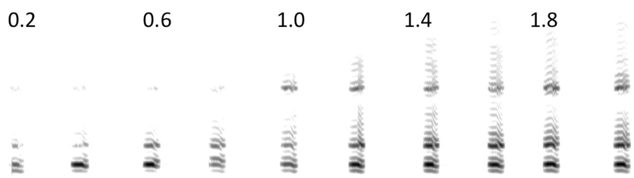


Figure 6: Increasing the glottal waveform parameter introduces higher frequency components

4.1.8 *Biological Factors*

For the semi-automatic performance mode (described in Section 4.2.2), it is necessary to include parameters that represent the laughing nature of the person. First, how quickly can the person recover upon being exposed to a laughter-inducing stimulus? If the person is slow to recover, he will likely laugh for a longer time. Second, how much intensity does it take for this person to vocalize laughter-notes? Clearly, some people break out into laughter more easily than others. These characteristics are specified by the decay rate and threshold parameters, respectively. See Figure 7 for a graphical illustration of these features.

4.2 Performance Modes

Three different modes of performance have been implemented on our laughter model.

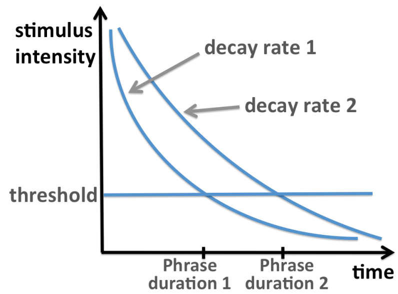


Figure 7: Modeling laughing-nature of an agent

4.2.1 *Manual*

The manual mode can be viewed as *an instrument* for laughter performance, allowing users to trigger and control laughter real-time using keyboad and trackpad controls. See Figure 8 for a summary of mappings.

4.2.2 *Semi-automatic*

The semi-automatic mode can be regarded as *an agent*, laughing according to its preset tendency and stimulus type. For this mode, the user supplies stimulus for laughter by either hitting or tilting the laptop, and the agent responds with dynamically synthesized laughter. Different stimulus types (e.g. a tickle, an evil thought, nervous energy, and sarcasm) are associated with different possible ranges of automated input parameters. Figure 9 shows a comparison of laughter triggered from four different stimulus types.

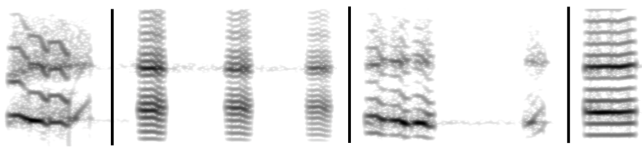


Figure 9: Laughter output from ticklish, evil, nervous, and sarcastic stimuli (left to right)

4.2.3 *Transcription*

The transcription mode serves as *a musical score.* This mode allows users to compose laughter by specifying the notes and phrases of laughter. In the current implementation, the score is written as a ChucK[2] [33] script that calls a function for each laugh-note to be synthesized, with control parameters supplied as function arguments.

5. SYSTEM

In this section, we briefly describe the implementation of our model. ChucK [33] offers a flexible environment for designing real-time audio synthesis, and combining it with SMELT [12] allowed us to easily leverage the laptop's keyboard, trackpad, and accelerometer-based motion sensor.

```
// sub patch
SndBuf buffy => TwoZero t => TwoZero t2 => OnePole p;

// unvoiced (noisy) component
Noise n => HPF highpass => t;

// formant filters
p => TwoPole f1 => Gain g;
p => TwoPole f2 => g;
p => TwoPole f3 => g;

// the rest
g => ADSR e => JCRev r => dac;
```

[2]http://chuck.stanford.edu/

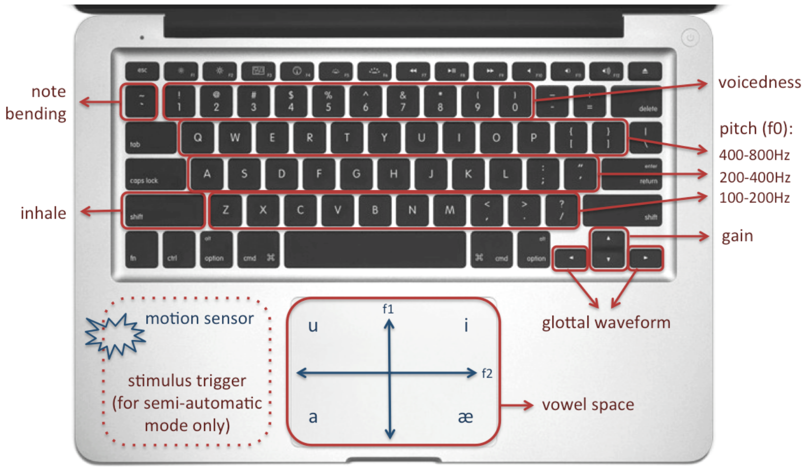


Figure 8: Control parameter mappings for real-time performance

As shown above, the synthesis patch used in our model is based on a formant-filter model for singing synthesis from motion-sing.ck[3]. A built-in sound buffer (`special:glot_pop`) from ChucK is used as the glottal source, and a Noise unit generator was added to introduce the noisy components.

The `bew()` function, taken from motion-sing.ck, implements dynamic formant manipulation: formants are generated by interpolating the frequencies of the formant filters based on the distances between the target-vowel and each of the four corners of a two-dimensional vowel space.

Modifications to the shape of the glottal waveform is achieved by changing the rate at which the buffer is read. Although this is a very rough simulation, changing the buffer rate effectively modifies the shape (slope) of the time-domain glottal waveform, consistent with how the glottis closes more abruptly in a pressed phonation.

An exponential decay function is used to model the behavior of a laughing agent (for the semi-automatic mode). The decay rate parameter is applied to the exponent, i.e. $f(x) = \exp(-decay\_rate \cdot kx)$, such that it controls laugh-phrase duration. The threshold parameter is applied as a cutoff such that if $f(x) < threshold$, then a note is no longer triggered, thereby terminating the laugh-phrase.

6. EVALUATION

We employed crowdsourcing using Amazon Mechanical Turk[4] to evaluate listeners' perception of our synthesized laughter. We prepared ten short laughter files to gauge their potential to convey social and emotional meaning, according to our operationalization of *expressiveness* from Section 2. For each laughter, we collected responses from five listeners, who were asked to describe what they believed to be a possible setting surrounding the laughter. Here we present sample responses to the ten laughter files, with the composer/ performer's original intention in parentheses:

1. (ticklish): "could not control", "laughs uncontrollably"
2. (like Tickle-Me-Elmo) "in the middle of a laugh panic"
3. (nervous high-pitch) "came across a result he did not expect", "heard some bad news... tries to laugh it off"
4. (nervous low-pitch): "uneasy, perhaps unsure of what he is laughing at", "maybe he is scared a bit"
5. (friendly): "laughed to impress the girl and give a positive reply for her talks"
6. (like Santa): "role playing Santa... in a mall with a queue of children", "Santa laughing in a calm way"
7. (sarcastic): "like a cartoon laugh... making fun of someone", "sarcastic laugh. unconvinced. teases"
8. (free): "playing with his brother", "play with infant"
9. (evil): "sinister laugh", "patronizing someone... insulting", "treating others as fool"
10. (contrived): "he is forced to laugh at something", "depressed", "an *act* of laughing"

These responses suggest that our model is capable of generating a range of laughter types and conveying *meaningful* expressions. Although some listeners found the synthesized laughter "weird", "scary", or "like a strange animal sound", others described it as "sweet", "unique", and even "sound[ing] like my laugh." Interestingly one listener commented, "possibility that the laugh is computer generated, not sure," implying that it was not aurally obvious that the laughter files had been fully synthesized.

7. APPLICATIONS

Our model, with its three modes of performance, naturally lends itself to a variety of applications. As a real-time controller with support for transcriptions, it can function as an expressive instrument for a laptop orchestra [28, 34]. For instance, we can compose a piece that features a virtuosic soloist bursting into a series of contagious laughter, and supporting ensemble members giggling back in response. On a more practical level, interactive laughter synthesis could be applied to speech synthesis, instilling a sense of emotional responsiveness to machine speech.

8. FUTURE WORK

Our current model is just a beginning in our efforts to synthesize laughter expressions. The model will naturally undergo iterative modifications to improve in usability and expressive potential; we outline two possible next steps here.

First, a usability study on the control interface would allow us to evaluate whether its design is consistent with the user's mental model. One possibility is to present synthesized laughter to subjects and instruct them to recreate the sound as closely as possible. Without prior instruction on how the keyboard and trackpad are mapped to the parameters, we may observe *discoverability* of our mappings. Alternatively, subjects can be given a quick tutorial on the instrument, and we may observe their usage to gain insights on the *learnability* of the controls.

[3] `http://smelt.cs.princeton.edu/code/motion/motion-sing.ck`

[4] `https://www.mturk.com/`. See [20] for techniques on using crowdsourcing for music perception experiments.

Another direction is to conduct a “challenge” for participants to compose as expressively diverse laughter as possible using our model. This challenge would serve as a creative way to understanding the expressive potential of our model. Based on the submissions, we can evaluate which types of laughter our system supports well, and which are more difficult to generate, motivating us to improve on weaknesses.

9. CONCLUSIONS

We now have an interactive tool, albeit a first prototype, in the form of a real-time controllable instrument that allows us to systematically play with the musical ingredients of laughter. We tried to define our model in terms of perceptually salient parameters in ways that prioritized expressiveness over realism. A demo of our system can be found at `http://vimeo.com/ge/lolol`. Future evaluation on the interface's usability and expressive potential should point us towards the next iteration.

10. REFERENCES

[1] B. Atal and S. Hanauer. Speech analysis and synthesis by linear prediction of the speech wave. *The Journal of the Acoustical Society of America*, 50(2B):637–655, 1971.

[2] J. Bachorowski, M. Smoski, and M. Owren. The acoustic features of human laughter. *The Journal of the Acoustical Society of America*, 110:1581, 2001.

[3] J. Chowning. Computer synthesis of the singing voice. In *Sound Generation in Winds, Strings, Computers*. Royal Swedish Academy of Music, 1980.

[4] J. Chowning. Frequency modulation synthesis of the singing voice. In *Current Directions in Computer Music Research*, pages 57–63. MIT Press, 1989.

[5] P. Cook. *Identification of Control Parameters in an Articulatory Vocal Tract Model, With Applications to the Synthesis of Singing*. PhD thesis, CCRMA, Stanford University, 1991.

[6] P. Cook. SPASM, a real-time vocal tract physical model controller; and singer, the companion software synthesis system. *Computer Music Journal*, 17(1):30–44, 1993.

[7] P. Cook and C. Leider. SqueezeVox: a new controller for vocal synthesis models. In *Proc. ICMC*, 2000.

[8] F. Cooper, A. Liberman, and J. Borst. The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings of the National Academy of Sciences of the United States of America*, 37(5):318, 1951.

[9] P. Delattre, A. Liberman, F. Cooper, and L. Gerstman. An experimental study of the acoustic determinants of vowel color; observations on one-and two-formant vowels synthesized from spectrographic patterns. *Word*, 1952.

[10] H. Dudley, R. Riesz, and S. Watkins. A synthetic speaker. *Journal of the Franklin Institute*, 227(6):739–764, 1939.

[11] T. Dutoit. *An introduction to text-to-speech synthesis*, volume 3. Springer, 1997.

[12] R. Fiebrink, G. Wang, and P. Cook. Don't forget the laptop: Using native input capabilities for expressive musical control. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 164–167, 2007.

[13] J. Flanagan. Voices of men and machines. *The Journal of the Acoustical Society of America*, 51(5A):1375–1387, 1972.

[14] D. Klatt. Software for a cascade/parallel formant synthesizer. *the Journal of the Acoustical Society of America*, 67:971, 1980.

[15] S. Kori. Perceptual dimensions of laughter and their acoustic correlates. *Proc. 11th International Congress of Phonetic Sciences*, 4:255–258, 1989.

[16] P. Lansky. Compositional applications of linear predictive coding. In *Current directions in computer music research*, pages 5–8. MIT Press, 1989.

[17] E. Lasarcyk and J. Trouvain. Imitating conversational laughter with an articulatory speech synthesizer. *Proceedings of the Interdisciplinary Workshop on the Phonetics of Laughter*, 2007.

[18] A. Liberman, P. Delattre, F. Cooper, and L. Gerstman. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied*, 68(8):1, 1954.

[19] M. Mathews. Bicycle built for two (1961). *Computer Music Currents no 13, The Historical CD of Digital Sound Synthesis*, 1961.

[20] J. Oh and G. Wang. Evaluating crowdsourcing through amazon mechanical turk as a technique for conducting music perception experiments. In *Proceedings of the 12th International Conference on Music Perception and Cognition*, 2012.

[21] J. Olive and M. Liberman. Text to speech – an overview. *The Journal of the Acoustical Society of America*, 78(S1):S6–S6, 1985.

[22] F. Poyatos. The many voices of laughter: A new audible-visual paralinguistic approach. *Semiotica*, 93(1-2):61–82, 1993.

[23] R. Provine. *Laughter: A scientific investigation*. Penguin Press, 2001.

[24] L. Rabiner. Digital-formant synthesizer for speech-synthesis studies. *The Journal of the Acoustical Society of America*, 43(4):822–828, 1968.

[25] X. Rodet. Time-domain formant-wave-function synthesis. *Computer Music Journal*, pages 9–14, 1984.

[26] T. Rossing, F. Moore, and P. Wheeler. *The science of sound*. Addison Wesley, San Francisco, CA, 2002.

[27] X. Serra and J. Smith III. Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, pages 12–24, 1990.

[28] S. Smallwood, D. Trueman, P. Cook, and G. Wang. Composing for laptop orchestra. *Computer Music Journal*, 32(1):9–25, 2008.

[29] S. Sundaram and S. Narayanan. Automatic acoustic synthesis of human-like laughter. *The Journal of the Acoustical Society of America*, 121:527, 2007.

[30] J. Sundberg. *The Science of the Singing Voice*. Northern Illinois Univ, 1987.

[31] D. Szameitat, K. Alter, A. Szameitat, D. Wildgruber, A. Sterr, and C. Darwin. Acoustic profiles of distinct emotional expressions in laughter. *The Journal of the Acoustical Society of America*, 126:354, 2009.

[32] D. Szameitat, C. Darwin, A. Szameitat, D. Wildgruber, and K. Alter. Formant characteristics of human laughter. *Journal of Voice*, 25(1):32–37, 2011.

[33] G. Wang. *The ChucK Audio Programming Language: A Strongly-timed and On-the-fly Environ/mentality*. PhD thesis, Princeton University, 2008.

[34] G. Wang, N. Bryan, J. Oh, and R. Hamilton. Stanford Laptop Orchestra (SLOrk). In *Proceedings of the International Computer Music Conference*, 2009.