PERCEPTION OF BEATING AND TWO-STAGE DECAY IN DUAL-POLARIZATION STRING MODELS

Hanna Järveläinen, Matti Karjalainen
Laboratory of Acoustics and Audio Signal Processing
Helsinki University of Technology
P.O. Box 3000, 02015 HUT, Espoo, Finland
hanna.jarvelainen@hut.fi, matti.karjalainen@hut.fi

Abstract: The perception of dual-polarization string sounds was studied in listening experiments using an acoustic guitar string model. Detecting a change in beating was measured as a function of the level difference between the vertical and horizontal components. Beating frequency and fundamental frequency were used as parameters. Perception of two-stage decay was studied in two experiments as a function of difference in the time constants of the polarization components. It was found that our sensitivity to dual-polarization effects is generally weak. Significant deviations from reference values were allowed before they became audible. The current results are useful in designing control schemes for synthesis parameters.

1 INTRODUCTION

The demand for high-quality audio in a low-bitrate channel has created a need for more parametric representations of sound. The MPEG-4 multimedia standard includes structured methods for representing synthetic audio (ISO/IEC, 1999), (Vercoe et al., 1998), (Scheirer and Yang, 2000). Also in the more recent MPEG-7 multimedia content description interface the timbre of musical sounds is described by perceptually relevant parameters (Peeters et al., 2000), (Lindsay and Kriechbaum, 1999). However, our understanding of the perception of musical sounds is not accurate enough for complete parametric representations. Perceptual studies are needed to explore the effects of individual attributes.

The objective of this study is to gain general understanding of musical sounds and to produce perceptual knowledge of the behavior of dual-polarization string models. However, a formal perceptual study is problematic because of the underlying complex beating and decay patterns that are highly multidimensional and thus difficult to explore. This first attempt concentrates on the general detection of the effects.

2 TWO-STAGE DECAY AND BEATS

After being excited a string vibrates freely in three modes: the transversal, the longitudinal, and the torsional. The last two modes have relatively little importance for sound production in most plucked string instruments. The transversal mode is divided into horizontal and vertical components, which vibrate in the plane of the top plate and the plane perpendicular to it, respectively. The vibrations attenuate exponentially, but because of unequal bridge impedance seen by the polarization components, they may have different decay times. The horizontal polarization is dominant at first, having a higher initial amplitude than the vertical component. However, it is decaying faster than the vertical component, which after a while becomes dominant. Thus the fast decaying but louder “prompt sound” is followed by the more sustained “aftersound”.

The unequal bridge impedance also causes a difference in the effective length of the string between the vertical and horizontal components. This results in a slight difference in the corresponding fundamental frequencies, which can be observed as beating, i.e., periodic amplitude modulation (AM) of the partials (Weinreich, 1977). Figure 1 gives an example of both phenomena. The left-hand panel shows the overall amplitude of a guitar sound that was synthesized by mistuning the polarization components while their decaying times were equal. The amplitude variations that are seen in the figure result from the complex effect of the beating patterns of individual harmonics. The right-hand panel presents the case where the components have equal fundamental frequencies but different decay times. The solid line shows clearly the two parts of the overall decay pattern. During the first second the sound is dominated by the horizontal component that is initially stronger but decays faster (dashed line). After that the more slowly decaying vertical component becomes stronger and creates the aftersound. The decay of the weaker component alone is shown by the dash-dotted line.

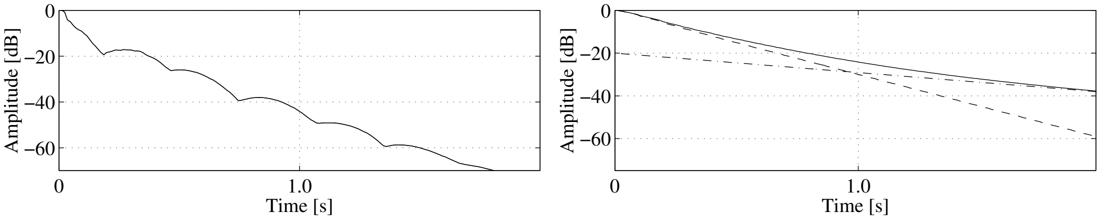

Fig. 1: Amplitude modulation (left) and two-stage decay (right) caused by double polarization.

2.1 PERCEPTION OF BEATS

The perception of beats in the audible range varies with modulation frequency. Two sinusoids at slightly different frequencies are perceived as a single beating tone whose beating frequency equals the frequency separation of the tones. Furthermore, if the amplitudes of the components of a beating tone are unequal, the pitch varies slightly at the speed of the modulations (Hartmann, 1997). The maximum of the pitch shift cycle coincides with the minimum of the amplitude, so that the pitch effect is supressed. However, reports exist on the fact that the pitch shift is sometimes detectable (Feth et al., 1982), (Versfeld and Houtsma, 1995).

The detection thresholds for amplitude modulation, expressed as modulation percentage required for detection, have been measured in previous studies as a function of modulation frequency (Schorer, 1986), (Moore and Sek, 1992), (Moore and Sek, 1995), being typically around 5 % and decreasing for modulation rates higher than 64 Hz.

3 THE DUAL-POLARIZATION STRING MODEL

The horizontal and vertical polarizations can be implemented by mixing the outputs of two basic string models (Jaffe and Smith, 1983). A dual-polarization string model of the acoustic guitar is presented in Fig. 2 (Välimäki et al., 1996), (Karjalainen et al., 1998). The input is fed to two parallel waveguide string models $S_h(z)$ and $S_v(z)$ according to the mixing parameter m_p. The outputs of the models are summed together. The beating effect is implemented by mistuning the delay lines corresponding to the horizontal and vertical modes, and the two-stage decay is controlled by varying the decay times of the string models. The parameter g_c between the components controls the coupling between the polarizations.

Fig. 3 presents the block diagram of a single string model (Karjalainen et al., 1998). The string has a transfer function

$$S(z) = \frac{1}{1 - z^{L_1} F(z) H_1(z)} \tag{1}$$

where L_1 is the integer part and $F(z)$ produces the fractional part of the delay line length. $H_1(z)$ is the loop filter which determines the decay of the tone according to

$$H_1(z) = \frac{g(1-a)}{1-az^{-}1} \tag{2}$$

where g controls the overall decay time constant and a is a frequency-dependent parameter. Two-stage decay is implemented by mistuning the g parameters in the horizontal and vertical string models.

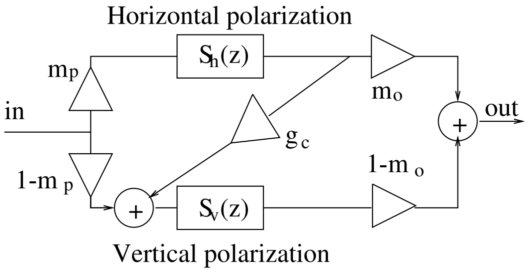

Fig. 2: The dual-polarization string model.

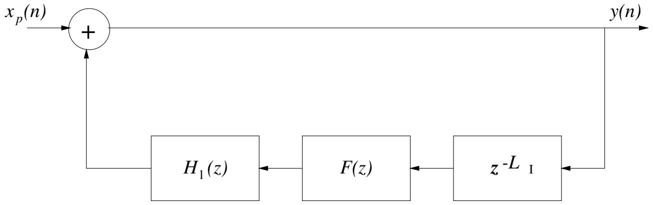

Fig. 3: Block diagram of a single string model.

4 LISTENING TESTS

The perception of dual-polarization sounds was studied in three listening experiments. Detecting a difference in beating was measured as a function of the level difference between the vertical and horizontal components. The second experiment studied the two-stage decay. Perceptual tolerances were measured for variations in the time constant of the aftersound. In the third experiment the perceived similarity of one- and two-stage decay patterns was measured.

4.1 TEST SOUNDS

The test sounds were synthetic acoustic guitar tones generated by the string model given above at sample rate of 44.1 kHz. The fundamental frequency f_0 and frequency separation Δf of the polarizations were used as parameters in both experiments. The tones were either E4 (330 Hz) or G3 (196 Hz) with Δf = 0, 0.4, or 0.7 cent (cent = 1/100 of a semitone). The largest mistuning was about 1 Hz for E4 and 0.6 Hz for G3. The coupling parameter g_c was set to zero and m_o to 0.5. Sound duration was 2.0 s in the beating experiment and 2.5 s in the decay experiment. The sounds ended with a fast linear fade-out. Longer sounds were used in the second experiment to let all sounds decay to the level of background noise so that the fade-outs could not be used as a cue in detecting two-stage decay.

4.2 TEST DESIGN AND SUBJECTS

In the first experiment the task was to detect a difference in beating when the level difference between the polarization components was varied. Measurements were made both starting at equal levels and decreasing the other component gradually, corresponding to increasing the mixing parameter m_p from 0.5, and starting from one polarization and increasing the level of the other component, which means in terms of synthesis parameters decreasing the value of m_p from 1.0. The measurement was run for E4 and G3 with $\Delta f = 0.4$ and 0.7 cents, which resulted in eight different test cases.

In the second experiment the decay time constants of the components were varied. The time constant of the faster decaying horizontal component was fixed to $\tau_h = 0.3015$ s, which corresponds to natural guitar tones. The time constant τ_v of the vertical component was increased from this level in seven steps. The two-stage decay became gradually audible as the difference in decay times increased. Fundamental frequency, mistuning of the components and their level difference were again used as parameters, but not in all combinations. Four cases were tested: E4 with 10-dB level difference between the components and either 0 or 0.7 cent difference in f_0, and G3 with tuned polarizations and either 10 dB or 20 dB level difference. The parameters are summarized in Table 1.

Experiments 1 and 2 followed the two-alternative forced choice (2AFC) paradigm. The seven conditions in each test case were repeated four times with the target signal occuring randomly in either the first or the second sound pair within the trial, and the task was to identify, which pair contained the target signal. The presentation order of the conditions was randomized within test cases. Detection thresholds were computed from the listening test data, expressed as stimulus intensity required for 76 % correct responses. Experiment 3 was a similarity rating test using a scale from 0 to 10. The judgments were recorded by precision of one decimal.

The tests were carried out in a quiet listening room one subject at a time. The sounds were played from a computer through Sennheiser HD 580 headphones at general level of about 85 dB, and the listeners gave their responses using the graphical interface of the GuineaPig2 system developed for listening experiments (Hynninen and Zacharov, 1999).

Seven subjects participated in the experiments. They were personnel of the HUT Laboratory of Acoustics and Audio Signal Prosessing, and most of them had previous experience of listening experiments as well as some musical background. None of them reported hearing defects. The subjects were trained before the experiment by a short session which presented examples of each test case.

Parameter	Beating experiment	Decay experiment
Note	E4 / G3	E4 / G3
Δ Level	Independent variable	-10 dB / -20 dB
Δf_0	0.4 / 0.7 cent	0 / 0.7 cent
τ_h	0.3015 s	0.3015 s
τ_v	0.3015 s	Independent variable

Table 1: Summary of parameters in experiments 1 and 2.

5 RESULTS

5.1 BEATING EXPERIMENT

The results of the beating experiment are presented in Fig. 4. The left-hand panel shows the detection thresholds when the amplitude of the vertical component was increased from zero, i.e., m_p was decreased from 1.0. The mean threshold over all cases was $m = 0.89$, which corresponds to a level difference of 18.2 dB between the polarizations. However, the ANOVA (analysis of variance) showed a significant effect of mistuning (P=0.05). The mean threshold for Δf = 0.7 cent was m_p= 0.91, while for Δf = 0.4 cent it was m_p = 0.86. The right-hand panel of Fig. 4 shows the thresholds when the components were equally strong at first and m_p was increased from 0.5. The mean threshold over all cases, m_p = 0.69 (7.0 dB), shows a large tolerance to changes. The effect of mistuning is again seen in the mean thresholds, being m_p = 0.64 for Δf = 0.7 cent and m_p = 0.72 for Δf = 0.4 cent. However, in this case the difference remains insignificant (P=0.12).

A tolerable decrease from m_p=1.0 was roughly 11 %, while an increase of 38 % was allowed from m_p=0.5. Thus smaller changes are detected around m_p = 1.0 than around m_p = 0.5. Just noticeable differences were unfortunately not measured for other values of m_p between 0.5 and 1. A general insensitivity to changes in the mid range would suggest that the effect of the mixing parameter would be almost two-valued: either ON or OFF. This would simplify its quantization. As can be seen from the boxplots in Fig. 4, there is not much space between the lowest results for m_p= 1.0 and the highest results for m_p = 0.5, and it is expected that our sensitivity in this area is not much better than in those that were measured.

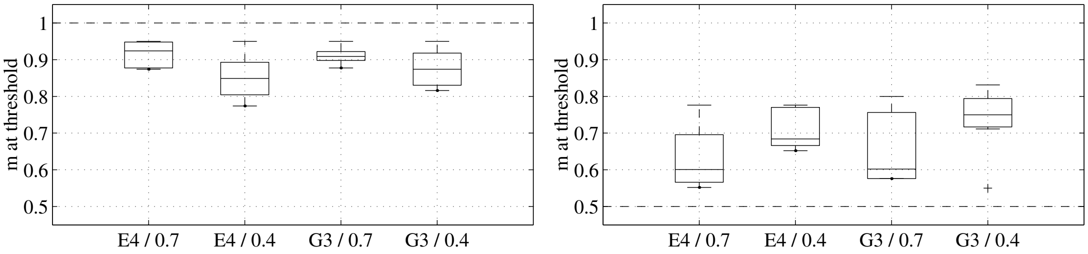

Fig. 4: Thresholds for detecting beats. Left: m_p decreasing from 1.0. Right: m_p increasing from 0.5.

5.2 DECAY EXPERIMENT

The increased time constants τ_v of the vertical component required for detecting two-stage decay are shown in Fig. 5. In the three leftmost cases the initial level difference between polarizations was 10 dB, and the analysis of variance (ANOVA) showed no significant difference between the results. The mean threshold over the 10-dB cases was τ_v = 0.43 s. In the fourth case the initial level difference was 20 dB, and the mean threshold was τ_v = 0.54 s. Compared to the reference value τ_h = 0.30 s, a difference of 43 % is tolerated for the first three cases and 80 % for the fourth case.

The thresholds are applied in synthesized sounds in Fig. 6. The solid lines in the left- and right-hand panels show the decay pattern of a G3 tone with initial level difference of 10 dB and 20 dB, respectively, synthesized by using the measured threshold values for τ_v. The dashed lines show the decay of the horizontal component, which has the same decay pattern as the reference tone with no two-stage decay, and the dotted line shows the decay of the vertical component. It is seen that the cross points, where the decay starts following the slower pattern, are roughly at t = 1.2 s and t = 1.3 s, and that the level difference at the end of the tone is around 15 dB. Expressed this way the differences between the cases are relatively small.

The motivation for the decay experiment was to find out, how easily a difference is perceived between two-stage decay tones and tones that decay like the horizontal component alone. The essential difference between these tones is the duration of the decaying part while the attack and early parts are practically equal. In the examples of Fig. 6 the −60-dB point is reached about 0.5 s later in the two-stage decay tone than in the reference tone until the difference becomes audible.

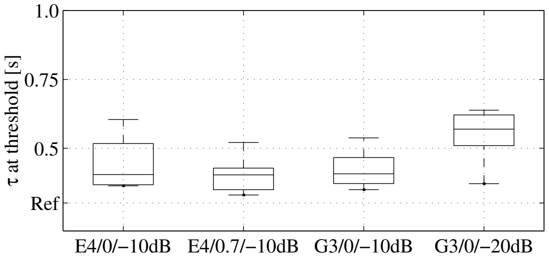

Fig. 5: Thresholds for detecting two-stage decay.

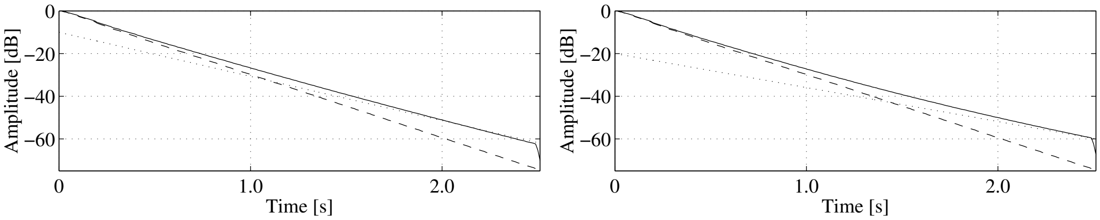

Fig. 6: The thresholds for detecting two-stage decay applied in synthesized sounds. Initial level difference of 10 dB (left) and 20 dB (right).

5.3 FORM OF DECAY ENVELOPE

Another approach to the perception of two-stage decay is to fix duration and measure the sensitivity to the notched form of the two-stage decay pattern. This was done by synthesizing exponentially decaying reference tones, whose amplitudes at the beginning and end were equal to those of a two-stage decay tone. The time constant of such sounds was between τ_h and τ_v of the polarization components.

The task in the third listening experiment was to judge the similarity of a two-stage decay tone and its corresponding reference tone on a scale from 0 to 10. The time constant τ_v of the vertical component was varied in five increasing steps between 0.54 s and 1.7 s, and the reference tones were computed accordingly. The measurement was run for G3 with 20-dB initial level difference between components. The stimuli were presented four times in randomized order. The results are seen in Fig. 7. The similarity ratings for the first two conditions are not significantly different, but for the third condition there is already a significant decrease (P=0.002). The last two conditions differ from the first condition even more.

The situation is depicted in Fig. 8 for case 3 (τ_v = 0.81 s) in the left-hand panel and case 5 (τ_v = 1.7 s) in the right-hand panel. The maximum level difference between the tones is 3.8 dB in the left panel and 6.5 dB in the right panel. According to the subjects, the difference was mainly detected as greater loudness in the mid segment of the one-stage decay tone.

6 DISCUSSION AND CONCLUSIONS

The perception of dual-polarization effects was studied in three listening tests. The detection of beats was measured as a function of level difference between the horizontal and vertical components, controlled through the mixing parameter m_p in the synthesis model. A decrease from m_p= 1.0 was detected more easily than an increase from m_p=0.5, for which a deviation of 38 % was allowed. A tolerable decrease from m_p = 1.0 was only 11 %. The region between 0.5 and 1.0 was not studied in this experiment, but it seems that accurate control of the level difference between the polarization

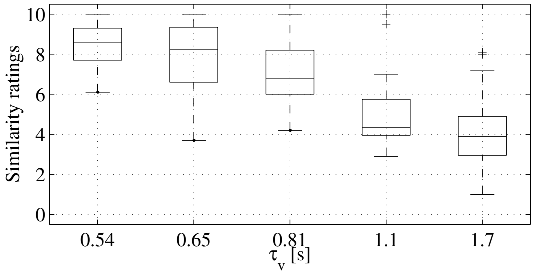

Fig. 7: Similarity ratings for one- and two-stage decay tones for $\tau_h = 0.30$ s and $\tau_v = 0.54 \ldots 1.7$ s.

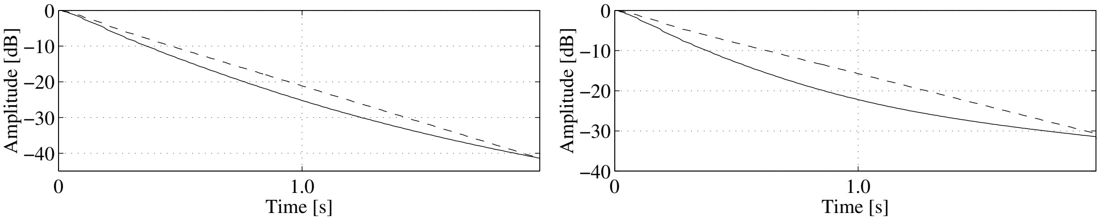

Fig. 8: Comparison of one-stage (dotted line) and two-stage (solid line) decay patterns. Left: $\tau_v = 0.81$ s (similarity rating 6.9/10), right: $\tau_v = 1.7$ s (similarity rating 4.0/10).

components is unnecessary.

Two-stage decay was studied in two experiments. At first its audibility was studied against sounds that decay like the horizontal component alone. This is similar to ignoring two-stage decay in sound synthesis. The detection mainly depended on the increased duration of the decaying part of the sound. Differences between 43 % and 80 % in the time constants of the horizontal and vertical components were allowed perceptually, depending on the level difference between the polarizations. In both cases, however, the cross point between the polarizations was roughly between $1.0 \ldots 1.3$ s at detection threshold.

Sensitivity to the notched form of the decay pattern was studied against exponentially decaying sounds, whose initial and final amplitudes matched those of the two-stage decay tones. This experiment showed that the form of the decay pattern is perceived as a loudness difference in the mid segment of the sounds. Two-stage decay is thus perceived in different ways depending on the reference tones. Compared to similarly synthesized sounds with only one polarization component, as in the first decay experiment, detection of two-stage decay is based on increased duration of the decaying part, while the notched form of the decay pattern is detected as a difference in loudness. The results of the first decay experiment are perhaps more relevant to sound synthesis.

The general sensitivity to dual-polarization effects is relatively weak. However, interaction of the beating and two-stage decay effects should be considered when designing control schemes for the synthesis models. If the polarization components are made equally strong, two-stage decay cannot be implemented at all. However, the notched form of the decay envelope is relatively easily discriminated from simple exponential decay, which suggests that two-stage decay has an effect on the perceived quality of synthetic tones.

The results are useful in control of the parameters in model-based synthesis. Whenever the effects of dual polarization remain inaudible, synthesis could be simplified by using only one string model. To fully apply perceptual knowledge in sound synthesis, the perceptual models should be used in parallel with efficient parameter estimation techniques. Much work is going on in parameter estimation; examples concerning the estimation and quantization of dual-polarization parameters are given in (Nackaerts et al., 2001) and (Riionheimo and Välimäki, 2002). However, the perception of beating and decaying effects in instrument sounds is interesting also in theoretical sense. More research is required for a thorough understanding of these phenomena.

ACKNOWLEDGMENTS

This work was supported by the Academy of Finland through the Pythagoras graduate school, and Nokia Research Center. The authors wish to thank the "voluntary" listeners.

REFERENCES

Feth, L., H. O'Malley, and J. Ramsey (**1982**), "Pitch of unresolved two-tone complex tones," *J. Acoust. Soc. Am.* **72**, pp. 1403–1412.

Hartmann, W. (**1997**), *Signals, Sound, and Sensation* (AIP Press, New York).

Hynninen, J. and N. Zacharov (**1999**), "GuineaPig – a generic subjective test system for multichannel audio," Presented at the Audio Engineering Society 106th Conv., May 1999, Munich, Germany, preprint no. 4871.

ISO/IEC (**1999**), "ISO/IEC IS 14496-3 Information Technology – Coding of Audiovisual Objects, Part 3: Audio," .

Jaffe, D. A. and J. Smith (**1983**), "Extensions of the Karplus-Strong plucked-string algorithm," *Computer Music J.* **7**(2), pp. 56–69.

Karjalainen, M., V. Välimäki, and T. Tolonen (**1998**), "Plucked-string models: from the Karplus-Strong algorithm to digital waveguides and beyond," *Computer Music J.* **22**(3), pp. 17–32.

Lindsay, A. and W. Kriechbaum (**1999**), "There's more than one way to hear it: Multiple representations of music in MPEG-7," *Journal of New Music Research* **28**(4).

Moore, B. and A. Sek (**1992**), "Detection of combined frequency and amplitude modulation," *J. Acoust. Soc. Am.* **92**(6), pp. 3119–3131.

Moore, B. and A. Sek (**1995**), "Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation)," *J. Acoust. Soc. Am.* **97**(4), pp. 2468–2478.

Nackaerts, A., B. De Moor, and R. Lauwereins (**2001**), "Parameter estimation for dual-polarization plucked string models," in *Proc. Int. Computer Music Conf.*, pp. 203–206, Havana, Cuba.

Peeters, G., S. McAdams, and P. Herrera (**2000**), "Instrument sound description in the context of MPEG-7," Proc. Int. Computer Music Conf., Berlin, Germany.

Riionheimo, J. and V. Välimäki (**2002**), "Parameter estimation of a plucked string synthesis model with a genetic algorithm," in *Proc. Int. Computer Music Conf. ICMC2002*, pp. 283–286, Gothenburg, Sweden.

Scheirer, E. D. and J.-W. Yang (**2000**), "Synthetic and SNHC audio in MPEG-4," *Signal Processing: Image Communication* **15**, pp. 445–461.

Schorer, E. (**1986**), "Critical modulation frequency based on detection of AM versus FM tones," *J. Acoust. Soc. Am.* **79**, pp. 1788–1803.

Välimäki, V., J. Huopaniemi, M. Karjalainen, and Z. Jánosy (**1996**), "Physical modeling of plucked string instruments with application to real-time sound synthesis," *J. Audio Eng. Soc.* **44**, pp. 331–353.

Vercoe, B., W. G. Gardner, and E. D. Scheirer (**1998**), "Structured audio: Creation, transmission, and rendering of parametric sound representations," *Proc. IEEE* **86**(5), pp. 922–940.

Versfeld, N. and A. Houtsma (**1995**), "Discrimination of changes in the spectral shape of two-tone complexes," *J. Acoust. Soc. Am.* **98**, pp. 807–816.

Weinreich, G. (**1977**), "Coupled piano strings," *J. Acoust. Soc. Am.* **62**, pp. 1474–1484.