be more useful because of the smaller amount of data available per class when there are multiple classes.

The problem of channel compensation, particularly in the model space, can be viewed as a special case of the more general problem of adaptation. Indeed, the ML feature-space method described here is similar to techniques used for speaker adaptation. For word recognition applications where more complex adaptation techniques are used, such as ML linear regression [13], channel compensation is implicit in the overall adaptation. Thus, the simple approaches described here are mainly appropriate for short utterances where more general adaptation is not feasible. A question raised by the connection between adaptation and channel compensation is whether it is possible to separately compensate for the channel and not the speaker for problems such as speaker identification or verification. We conjecture that by having separate priors for the channel and the speaker, the Bayesian approach will help solve channel and speaker separation.

REFERENCES


[10] F. Scalcon is with Generalmusic S.p.A., 47842 Rimini, Italy (e-mail: rocchesso@sci.univr.it). D. Rocchesso is with the Department of Science and Technology, University of Verona, 37134 Verona, Italy (e-mail: rocchesso@sci.univr.it).

Bandwidth of Perceived Inharmonicity for Physical Modeling of Dispersive Strings

Davide Rocchesso and Francesco Scalcon

Abstract—The influence of accurate reproduction of inharmonicity on the perceived quality of piano tones is investigated. Acoustic piano tones were resynthesized by changing the bandwidth of correct positioning of partials. Cutoff frequencies of inharmonicity for different pitches have been established by psychoacoustic experimentation. The results are applicable to the design of dispersive resonators in sound synthesis by physical modeling.

Index Terms—Allpass filters, inharmonicity, physical modeling, psychoacoustics, wave dispersion.

I. INTRODUCTION

The strings of acoustic pianos are not perfectly flexible and, as a consequence, velocity waves travel without preserving their shape. In other words, string stiffness implies wave dispersion. Moreover, dispersion implies inharmonicity of piano tones. While the physical behavior of stiff strings is well understood [1], [2], we are not aware of much literature on the perceived effects of inharmonicity on the quality of piano tones. Podlesak and Lee pointed out the importance of pitch glides due to inharmonicity of low bass piano tones such as the “A0” [3] (fundamental, 27.5 Hz). These glides have a duration of a few tens of milliseconds, and they can be discriminated by the listener, especially during the initial transient. Our experience has shown that inharmonicity also plays a very important role during the sound decay stage, even for notes much higher than “A0” whose glides are too short to be discriminated. Experienced musicians often report that synthetic sounds with harmonic structure are too “static” during decay. These informal findings are supported by the results obtained by Moore et al. [4], who showed that more than a second of listening is needed to fully appreciate the timbral effect of a mistuned partial in an otherwise harmonic complex tone.

The main question we want to raise is “How accurate does the alignment of partials of a synthetic piano tone have to be in order that it may be perceived as natural?” To perform some experiments, the question has to be reformulated in a more strict form, according to our goals. After many discussions with pianists, we realized that the main problem with synthetic piano tones that do not follow the natural inharmonic distribution of partials is a lack of appropriate liveliness in the decay stage. Therefore, we decided to focus on the decay of piano tones. Since our curiosity into the aural perception of inharmonicity arose after intensive experimentation on physical modeling of the piano [6], we restricted the field of investigation even further, by taking the needs of modeling into account. In physical modeling of strings it is customary to lump losses and dispersion, and simulate them by means of a couple of linear filters [7]. Our experience has

Manuscript received October 23, 1998; revised January 12, 1999. This research was conducted under the auspices of Generalmusic S.p.A. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dennis R. Morgan.

D. Rocchesso is with the Department of Science and Technology, University of Verona, 37134 Verona, Italy (e-mail: rocchesso@sci.univr.it).

F. Scalcon is with Generalmusic S.p.A., 47842 Rimini, Italy (e-mail: francescos@generalmusic.com).

Publisher Item Identifier S 1063-6676(99)06565-7.

1 According to Hartmann et al. [5, p. 1716], the detection of inharmonicity in [4] was based upon beats and roughness.
shown that the allpass filter used for dispersion simulation has to be of high order if the dispersion curve of low-pitched notes is to be approximated with good accuracy [8]. Namely, the allpass filters for the first octave can have order 16 or more. High-order allpass filters are difficult to design and expensive to implement, and so we started looking for psychoacoustically driven design criteria. The first fundamental question was whether there is a preferred frequency range for dispersion approximation, i.e., a cutoff frequency above which it doesn’t make sense to approximate the correct positioning of partials. The present correspondence deals with this question and gives an answer motivated by a psychoacoustic experiment.

We remark that the experiment we are presenting aims at achieving a more efficient filter design rather than understanding sound perception. In fact, for synthesizing dispersive strings we proposed using allpass filters designed by means of a modification of the Lang–Laakso technique [8], [9], which provides a weighted least-squares phase error approximation over a discrete set of frequencies. As a viable design set we use the frequency values of the (inharmonic) partials, and we are interested in making this set small in order to reduce the complexity of both the design procedure and the designed filters [8].

The perception of inharmonicity has received some attention in the literature of psychoacoustics [4], [5], [10]. In particular, thresholds for the segregation of a mistuned partial in an otherwise harmonic complex tone have been measured as a function of partial frequency and tone duration [5]. Even though these works give useful hints to the sound-processing designer, it is difficult to translate results obtained by means of very artificial sound stimuli into prescriptions for filtering actual musical sounds. The need of psychoacoustic experiments such as the one we are presenting has emerged from practice in physical modeling for sound synthesis [7, p. 453]. More generally, any sound synthesis techniques which work by exciting a linear, possibly inharmonic resonator [11] can benefit from the practice in physical modeling for sound synthesis [7, p. 453]. More specifically, the present correspondence deals with this question and provides an answer motivated by psychoacoustic experiments.

II. Experiment

A. Stimuli

The stimuli were produced by means of additive synthesis, in such a way that detailed control on the position of each partial is possible. Additive synthesis was driven by results of analysis of the sounds produced by an acoustic piano Schulze–Pollmann 190-F. Namely, the notes "C1," "C2," "C3," "C4," and "C5" were played forttissimo in an anechoic chamber and recorded with a microphone positioned where the head of the player would be during normal playing conditions. The analysis was performed using the short-time Fourier transform and isolating the partials from the "stochastic" part of sounds [12]. Instantaneous frequency and amplitude data for all significant partials were extracted and stored for use during resynthesis. Table I reports the analysis parameters for the five notes.

<table>
<thead>
<tr>
<th>Note</th>
<th>Sampling Rate</th>
<th>Window Type</th>
<th>Window Size</th>
<th>Hop Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>16742 Hz</td>
<td>Kaiser</td>
<td>4096</td>
<td>1024</td>
</tr>
<tr>
<td>C2</td>
<td>24000 Hz</td>
<td>Kaiser</td>
<td>4096</td>
<td>1024</td>
</tr>
<tr>
<td>C3</td>
<td>48000 Hz</td>
<td>Kaiser</td>
<td>4096</td>
<td>1024</td>
</tr>
<tr>
<td>C4</td>
<td>48000 Hz</td>
<td>Kaiser</td>
<td>2048</td>
<td>512</td>
</tr>
<tr>
<td>C5</td>
<td>48000 Hz</td>
<td>Kaiser</td>
<td>1024</td>
<td>256</td>
</tr>
</tbody>
</table>

Table I

Instantaneous frequencies were estimated by computing the backward difference of the transform phase between neighboring notes [13]. In order to improve the frequency resolution for "C2" and "C1" while keeping short time windows (4096 points), we had to reduce the sampling rate from the nominal $F_s = 48000$ to 24000 and 16742 Hz, respectively.

Five sets of tones were synthesized.

- $Cn_{ADD}$: Resynthesis without modification of the extracted data.
- $Cn_{NFM}$: Resynthesis after elimination of frequency fluctuations of partials.
- $Cn_{F}$: Resynthesis without frequency fluctuations and with partial frequency positions set according to the theoretical curve, computed using an inharmonicity coefficient [2] derived from measured physical properties of the strings, reported in Table II.
- $Cn_{E}$: Resynthesis without frequency fluctuations and with partial frequency positions set according to the theoretical curve. The inharmonicity coefficient has been chosen qualitatively in such a way that the theoretical curve follows the distribution of experimental partial frequencies.
- $Cn_{J}$: Resynthesis equal to $Cn_{E}$ up to a cutoff frequency $f$. Residual partials higher than $f$ are positioned at points regularly spaced in frequency.

The only exception to these rules is found for the note "C1," where a good fit of rather dispersed data was found to be given by the physics-driven theoretical positioning, and the resulting tone was named $C1_{T}$. Two other tones were synthesized using an inharmonicity coefficient increased by 10% and 20%, respectively.

As an example, Fig. 1(a) shows the measured dispersion (i.e., deviation from harmonic position expressed as a percentage) for partials of the note "C2" and two theoretical curves obtained by using the inharmonicity coefficient derived from physical data and from qualitative point fitting, respectively. Fig. 1(b) shows the dispersion of partials of resynthesized tones $C2_J$; the cutoff points are clearly visible as the partial number where the dotted curves stop increasing. The cutoff frequencies $f$ were spaced according to critical bands, at intervals of 2 Barks [14]. A potential bias for the experiment might occur if no limit is placed on the energy bandwidth of the synthesized tone. In fact, differently dispersed versions of the same note with the same number of partials can have different bandwidths, and this phenomenon can deceive the listener who might rank two tones differently simply because they have a different brightness. Therefore, a constant upper limit represented by the dashed line in Fig. 1(b)] was enforced on the resynthesized partials, and no partial having frequency higher than that limit was reproduced.

B. Method

The tones described in Section II-A were synthesized and digitally recorded on a digital audio tape. All the subjects were requested to listen to the tape by means of headphones.

1) Pretest: A preliminary test was conducted in order to ascertain the accuracy of the synthesized tones in reproducing the perceived effects of inharmonicity. For each note, the recorded tone and the three stimuli $Cn_{ADD}$, $Cn_{NFM}$, and $Cn_{E}$ were played in sequence. These tones represent increasing degrees of simplification in the reproduction of the acoustic piano tones. The subjects were requested to focus on the decay stage and to describe any differences in timbre. The answers confirmed that the simplifications do not alter the perceived effects of inharmonicity, thus indicating that the liveliness in the decay stage is neither due to rapid fluctuations nor due to the microstructure of the dispersion curve. Another pretest was needed to...
TABLE II

<table>
<thead>
<tr>
<th>Note</th>
<th>Length L [mm]</th>
<th>String Diameter [mm]</th>
<th>Steel Section S [mm²]</th>
<th>Tension T [N]</th>
<th>Young Mod. Y [N/m²]</th>
<th>Inhm. Coeff. β = πY S² / 4LU²T</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>1390</td>
<td>5.250</td>
<td>1.431</td>
<td>1259.2</td>
<td>1.95E11</td>
<td>0.00012898</td>
</tr>
<tr>
<td>C2</td>
<td>1232</td>
<td>2.560</td>
<td>0.908</td>
<td>952.6</td>
<td>1.95E11</td>
<td>0.0008725</td>
</tr>
<tr>
<td>C3</td>
<td>1187</td>
<td>1.075</td>
<td>0.908</td>
<td>687.1</td>
<td>1.95E11</td>
<td>0.0013032</td>
</tr>
<tr>
<td>C4</td>
<td>649</td>
<td>1.025</td>
<td>0.825</td>
<td>747.0</td>
<td>1.95E11</td>
<td>0.0030144</td>
</tr>
<tr>
<td>C5</td>
<td>338</td>
<td>0.975</td>
<td>0.747</td>
<td>733.3</td>
<td>1.95E11</td>
<td>0.00101949</td>
</tr>
</tbody>
</table>

Fig. 1. (a) Measured dispersion expressed as percentage for partials of the note “C2” and two theoretical curves obtained by using the inharmonicity coefficient derived from physical data and from point fitting, respectively. (b) Dispersion of partials of resynthesized tones C’2; dispersion cutoff frequencies are indicated.

ascertain whether the choice of the inharmonicity coefficient is critical or not. To this end, the stimuli CnI and CnE were played one right after the other, and the subjects had to describe perceived differences in the decay stage of sound. Answers indicated that there are only subtle perceived differences between the different dispersion curves, and that the choice of the inharmonicity coefficient is not critical for a correct reproduction of a lively decay. This is confirmed by the fact that pianos of good quality are largely different in terms of materials and size, thus showing significant differences in string inharmonicity.

2) Main Experiment: The main experiment was conducted by playing the tones CnI in sequence, using a randomly chosen permutation of cutoff frequencies I. The subjects had to locate each tone on a (linear) scale of perceived naturalness, ranging from good/natural to unacceptable/unnatural. This operation consisted in drawing a mark along a 61 mm-long segment. The subjects were advised to focus on the decay stage of sounds and to neglect possible differences in the attacks. These subtle differences were difficult to eliminate as they are a byproduct of resynthesis. Actually, two subjects had to be discarded after realizing that they misunderstood the focus of the experiment. The chosen subjects are all Italian-speaking musicians accustomed to analyzing timbre in all its multiple facets, and with no reported hearing damage. For each note, each subject had to evaluate all the resynthesis using different cutoff frequencies. Eight subjects were used for all the notes except for “C1” which was synthesized and tested a few weeks later. Only a subset of six subjects could be used for note “C1.”

C. Results

All the subjects reported that the task of the main experiment was not easy. Nevertheless, they said that it is possible to rank the sound with a good degree of certainty. In Figs. 2–4 we report the mean ratings over all subjects and the standard deviations for the various cutoff frequencies of every note. An evaluation close to zero indicates a sound decay perceived as natural, while a value close to 61 indicates an unnatural sound decay. All the notes, except the “C5,” give a clear indication of an increasing mean perceived quality versus increasing cutoff frequency.

1) Analysis of Variance: The reliability of the reported evidence is confirmed by the analysis of variance (ANOVA) [15], giving the statistics reported in Table III. The column of p values (or significance levels) shows that the null hypothesis (i.e., the cutoff frequency doesn’t affect the perceived quality of timbre) is fulfilled with very small probability for the notes “C2” and “C3.” For notes “C1” and “C4” the null hypothesis is fulfilled with about 20% probability and for note “C5” this hypothesis is quite likely, thus indicating that reported behavior of “C5” does not show a statistically significant dependence on cutoff frequency. The note “C1” was the most difficult to analyze and resynthesize, due to its low pitch, and the group of subjects was smaller than for the other notes.

2) Bandwidth of Perceived Inharmonicity: The experiment confirms a phenomenon that has been intuitively known for a long time:

However, no audiometry was performed before the actual listening test.
the effect of inharmonicity is mainly important over the lower half of the keyboard. For notes such as “C5,” the upper partials only contribute to a brightness effect which is largely unaffected by their positions. Figs. 2–4 depict the graphs of measured mean ratings versus cutoff frequency. For the tests where the ANOVA gives a probability of rejection of the null hypothesis that is less than 20%, the curves are monotonic, thus indicating that better results are obtained for higher numbers of accurately positioned partials. Since we have measured the quality of piano sounds using a scale that ranges from unnatural to natural, we can set an arbitrary threshold of acceptability between the two extremes. If we set this threshold at the middle of the vertical scale, the frequency where the curves cross the middle horizontal line can be chosen as a bound of correct positioning of partials. In other words, when synthesizing pianolike tones, we have to position with good accuracy at least the partials which are below that point. Even though the bandwidth of perceived inharmonicity is smaller for lower notes, many more partials have to be positioned properly for low bass notes. For example, according to our experiment for the note “C1,” having about 1700 Hz as the bandwidth of perceived inharmonicity, about 50 partial frequencies have to be approximated. On the other hand, for the note “C3” about 30 partials will suffice.

The results of this experiment are valuable for designing the allpass filters needed by physical models of dispersive resonators. The effectiveness of design is significantly improved if the design procedure refines its approximations in the frequency areas where high precision is needed.

Of course, the accuracy in the bandwidth of perceived inharmonicity can be ensured in several ways, for example by minimizing the maximum error on partial positions or by minimizing the sum of squared errors. The best error criterion for allpass filter design remains to be found by other psychoacoustic experiments.
TABLE III
ANALYSIS OF VARIANCE FOR THE PERCEIVED QUALITY OF RESYNTHESIZED TONES

<table>
<thead>
<tr>
<th>Note</th>
<th>df</th>
<th>F-ratio</th>
<th>p-value</th>
<th>crit. F-ratio (0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>4</td>
<td>1.7523</td>
<td>0.1702</td>
<td>2.7587</td>
</tr>
<tr>
<td>C2</td>
<td>4</td>
<td>10.1846</td>
<td>0.0000</td>
<td>2.6415</td>
</tr>
<tr>
<td>C3</td>
<td>4</td>
<td>10.4444</td>
<td>0.0000</td>
<td>2.6415</td>
</tr>
<tr>
<td>C4</td>
<td>5</td>
<td>1.4941</td>
<td>0.2123</td>
<td>2.4377</td>
</tr>
<tr>
<td>C5</td>
<td>5</td>
<td>0.9967</td>
<td>0.4315</td>
<td>2.4377</td>
</tr>
</tbody>
</table>

Fig. 5. Bandwidth of perceived inharmonicity as a function of fundamental frequency.

III. CONCLUSION

The main result of this experiment can be summarized as a sparse set of points giving the bandwidth of perceived inharmonicity of piano tones as a function of fundamental frequency. Fig. 5 displays these findings, which can be used to drive the allpass filter design procedures toward the lowpass frequency ranges of higher significance. The three points of Fig. 5 were drawn on the basis of frequencies where the corresponding curves of Figs. 2 and 3 cross the line of middle vertical scale. For the note “C4” a corresponding point could not be drawn, since the corresponding curve of Fig. 3 was all below the middle of the vertical scale and the increase in quality was much more gradual than in the other cases. The bandwidth of perceived inharmonicity seems to increase roughly linearly when pitch varies along a log scale. While for low pitches (<70 Hz) the bandwidth of perceived inharmonicity for piano tones is below the frequency limit for the segregation of a mistuned harmonic [5], for higher pitches it extends above that limit, thus confirming that segregation and detection of timbral effects are related yet different tasks.

Further experiments have to be performed in order to understand what kind of norm of the error on partial positions has to be minimized in order to minimize differences in perceived timbre.

ACKNOWLEDGMENT

G. Borin gave useful hints for conducting the experiment and helped applying the results to digital piano design.

REFERENCES


A Note on “A Secondary Path Modeling Technique for Active Noise Control Systems”

Dennis R. Morgan

Abstract—A paradox arises in a recent paper by Kuo and Vijayan because notching out narrowband primary noise would seem to also prevent secondary path identification at the primary center frequency. This conundrum is resolved by examining the notch bandwidth in comparison to the resolution of the model filter.

Index Terms—Active noise control, adaptive filters, adaptive signal processing, adaptive systems, modeling.

In [1], a pseudorandom probe signal is used in conjunction with an active noise control (ANC) system for identification of the secondary path. Further experiments have to be performed in order to understand what kind of norm of the error on partial positions has to be minimized in order to minimize differences in perceived timbre.

Acknowledgment

G. Borin gave useful hints for conducting the experiment and helped applying the results to digital piano design.

IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 7, NO. 5, SEPTEMBER 1999 601

Publisher Item Identifier S 1063-6676(99)-06563-3.