Decomposition of Steady State Instrument Data into Excitation System and Formant Filter Components

Ilia Bisnovatyi and Michael O'Donnell
Department of Computer Science, University of Chicago
1100 E. 58th street, Chicago IL 60637, USA
ilia@cs.uchicago.edu, odonnell@cs.uchicago.edu

Abstract

This paper describes a method for decomposing steady-state instrument data into excitation and formant filter components. The input data, taken from several series of recordings of acoustical instruments is analyzed in the frequency domain, and for each series a model is built, which most accurately represents the data as a source-filter system. The source part is taken to be a harmonic excitation system with frequency-invariant magnitudes, and the filter part is considered to be responsible for all spectral inhomogenieties. This method has been applied to the SHARC database of steady state instrument data to create source-filter models for a large number of acoustical instruments. Subsequent use of such models can have a wide variety of applications, including wavetable and physical modeling synthesis, high quality pitch shifting, and creation of "hybrid" instrument timbres.

1 Introduction

Currently, two techniques are used most successfully for digital emulation of acoustical instruments – physical modeling and sampling (or wavetable synthesis). Both have their inherent limitations – physical modeling is, in general, computationally expensive and requires a good understanding of the mechanical workings of a particular instrument, whereas sampling is poorly informed of the physical properties of instruments and is not well-suited for modeling transitions or dynamic control of sound. In addition, sampling often relies on pitch shifting by varying the playback speed of a sample – a technique that disregards the spectral "signature" of an instrument and often creates undesirable audible artifacts.

In order to overcome these limitations, various intermediate representations have been sought. A source-filter model, well-known from speech research [1] has been applied to both physical modeling [3] and sampling [4] and proved useful for providing better control of timbre and reducing pitch shifting artifacts. When combined with well-designed representations for the excitation system, it can lead to robust synthesis methods.

This paper describes an algorithm, which allows to obtain magnitude data for the formant filter and the excitation part of the source-filter model, given a series of steady state data from a particular harmonic instrument (or voice). The algorithm treats the decomposition as a fitting problem, finding a solution that most accurately represents the original data. In general, the problem under-determined; by selecting the right constraints, one should be able to obtain a solution that will be most appropriate for further applications.

2 Input Data

While the algorithm described in this paper can be applied to any set of steady-state data, we tested it out on SHARC, a timbral database covering a large number of acoustical instruments, which is free and is readily available on the web [2]. For each instrument, a chromatic series of notes has been analyzed (every note had been individually played and digitally recorded; a description of the original recordings can be currently found in [6]). Several periods of steady-state sound had been selected from each note in a series and spectrally analyzed. SHARC assumes that all input sounds are harmonic, and since for every note the fundamental is known, the steady-state data can be represented as a set of values for magnitudes and phases of partials. The total number of detected partials varies from note to note, and the range and total number of notes varies from instrument to instrument. For the purposes of this paper, only the magnitude data is utilized; phase information can be added later to get a more precise description of the formant filter.

3 Representation

Let S be the total number of chromatic samples in a series, and K – the smallest number of available partials for any given sample (for the purposes of uniformity, we choose to consider the same number of partials for every sample; higher notes will tend to have fewer partials due to the Nyquist limit, therefore one can either disregard some of the partials for the lower notes, or truncate the input series). Let $D_{i,j}$ be the amplitude of j-th partial of i-th note – these are the data points. Now consider an equally spaced grid in the log frequency space, whose bins are centered on the fundamentals of equally tempered chromatic tones. This grid will define the resolution for the formant filter coefficients R_n, i.e. for each bin the magnitude of filter's frequency response in that bin will have to be determined. This resolution is reasonable, because the formant curve is expected to be fairly smooth and because for most traditional applications one will rarely need to synthesize notes less than a semitone apart (however, if required, an interpolated curve can be used). Note also that this resolution is only determined by the spacing of the original samples, and adapting to a more finely sampled input would be trivial. The target excitation system will consist of K partials with amplitudes P_j, which remain constant for every sample. Additionally, to account for the differences in the musical performance of individual notes, an overall multiplicative scaling coefficient A_i for each sample is introduced.

The data points and the variables are related by a set of equations

$$D_{i,j} = A_i P_j R_n \qquad (1)$$

for i=1..S and j=1..K. Index n is the number of the bin into which the frequency of the j-th partial of i-th sample falls, starting with 1 for the fundamental of the lowest note, i.e. $n = \lfloor 12 \log_2 j + 1/2 \rfloor + i$.

All of the values in (1) are positive, and thus, to facilitate the solution, the products in (1) can be easily converted into sums by switching to a logarithmic magnitude scale:

$$d_{i,j} = a_i + \mathrm{p_j} + \mathrm{r_n} \qquad (2)$$

where $d_{i,j} = \ln(D_{i,j})$, $a_i = \ln(A_i)$, $p_j = \ln(P_j)$, and $r_n = \ln(R_n)$. This is a system of $S*K$ linear algebraic equations; the data matrix $\{d_{i,j}\}$ can be collapsed into a vector $\overline{d}$, and all the variables – into a vector $\overline{v}$, thus transforming (2) into a linear system

$$\overline{v}\,\mathrm{M} = \overline{d}\,, \qquad v_k = \begin{cases} p_k, & k \le K \\ a_{k-K}, & K < k \le S+K \\ r_{k-S-K}, & k > S+K \end{cases} \qquad (3)$$

where M is the corresponding matrix of zeroes and ones. *Figure 1* shows a graphical representation of M for S=12, and K=16, with ones marked in black :

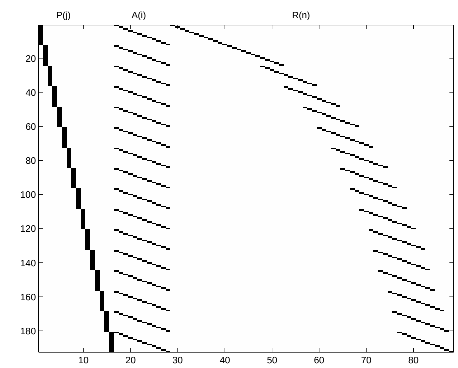

Fig.1

4 Approaches to the solution

The system (3) is generally underdetermined, since the rank of M is always less than $S+K+N$. One extra degree of freedom can be easily eliminated – an overall scaling factor that could be applied to the excitation at the expense of scaling coefficients A_i. However, even after normalizing the excitation (setting p_0=0 and eliminating the first column from M), the system will remain underdetermined (for all practically interesting cases this can be verified empirically by computing the rank of M).

There are many ways in which constraints could be added to (3) in order to choose the solution. For example, assumptions could be made about the smoothness of the filter or about the range into which the scaling coefficients $\{A_i\}$ fall. For the general case, after some experimentation, a robust iterative method was chosen. The iterations alternate between solving for $\{p_j\}$ given $\{r_n\}$ and solving for $\{r_n\}$ given [normalized] $\{p_j\}$. No special assumption is made about the values of a_i – they are readjusted after each iteration. A weighted least-squares convergence metric is used as a test for the termination of the iterative process. For every instrument from SHARC this algorithm converges within 20 iterations, allowing for deviations of $<0.1\%$.

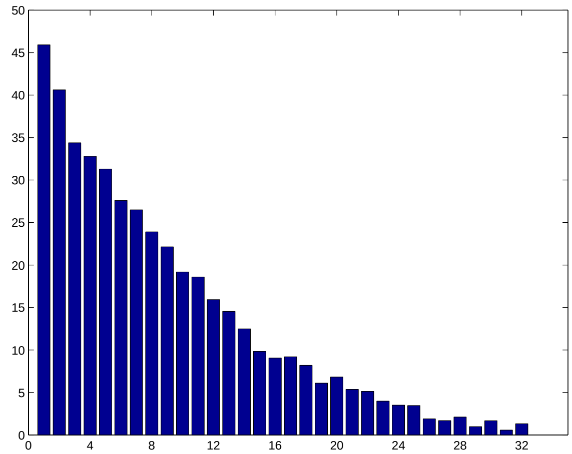

Fig.2

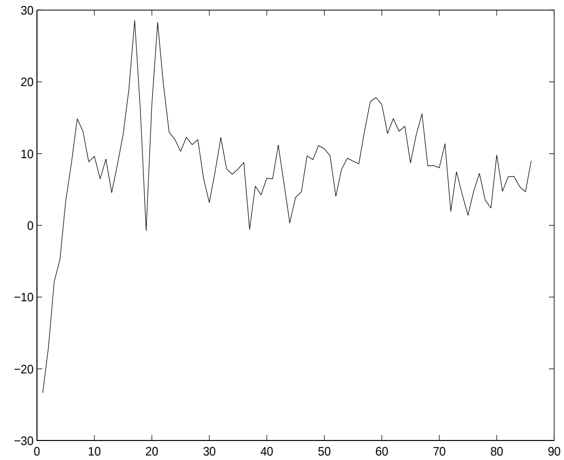

Fig.3

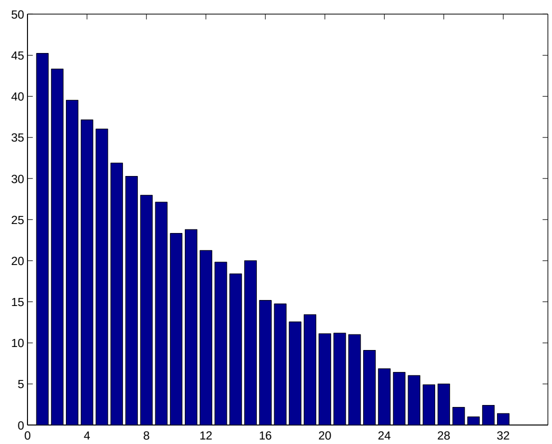

Fig.4

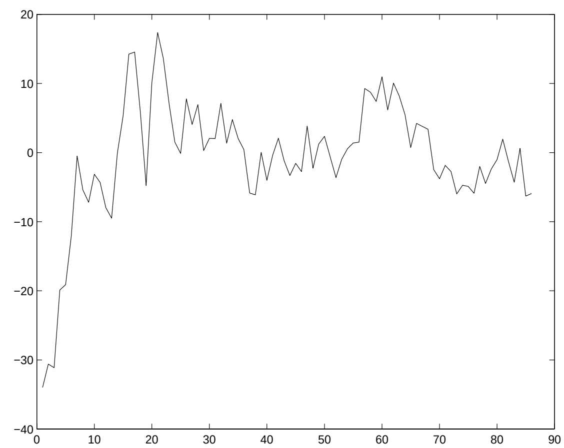

Fig.5

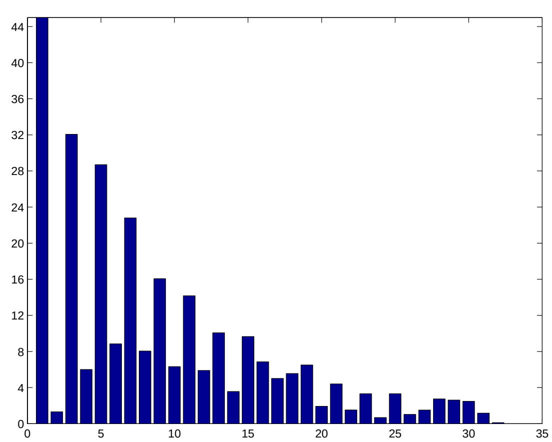

Fig.6

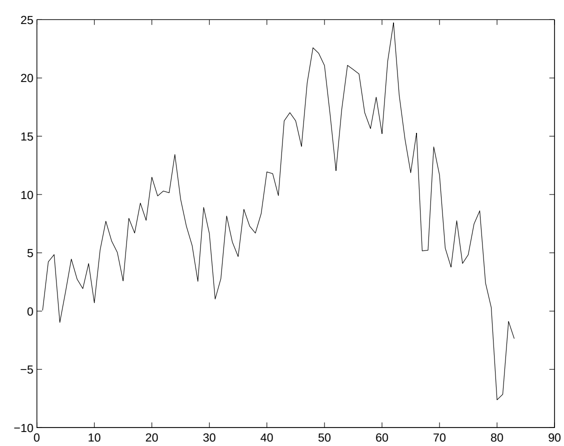

Fig.7

5 Results

The following figures illustrate some of the strengths and shortcomings of the proposed method. The pairs of figures 2,3 and 4,5 show the excitation and filter solutions for plucked and bowed cello respectively (S=28, K=32, F_1=65.4060Hz). As one would expect, the excitations are somewhat different, while the filter curves exhibit similar resonant properties, although they are not identical. Figures 6 and 7 show the excitation and filter solutions for bass clarinet (S=25, K=32, F_1=69.2960Hz). The suppression of even partials is clearly evident in the excitation, which conforms to the physical process of harmonic generation in clarinets [5].

6 Summary and discussion

The algorithm described here provides a fast and simple tool for obtaining excitation and filter components from steady-state magnitude data. The problem is reduced to a system of linear equations, which is generally under-constrained, and an iterative solution method has been proposed, which, we believe, selects qualitatively appropriate solutions. The final representation of the original magnitude data is precise; there is no data loss. An automated interface for the SHARC database has been built, providing excitation and filter patterns for a large number of acoustical instruments.

There are several directions for further improvement. As was mentioned previously, applying carefully chosen constraints to the variables can lead to a more appropriate choice of solution. With a slight modification, the algorithm could collect more information in cases when different excitation patterns are processed by the same filter (such as recordings of the same instrument played via different techniques) or the same excitation applied to different filters (e.g. a voice singing different vowels). Similarly, more information can be obtained by analyzing the same series played a number of times, since repetition will tend to average out the effects of uneven performances. Phase information can be added to obtain a more complete filter representation. Finally, once filter's magnitude and phase response is established, it could be used to determine the behavior of partials in attacks and other non-steady-state portions of the sound.

References

[1] T.F. Quartieri, R.J. McAulay, "Shape Invariant Time-scale and Pitch modification of Speech," *IEEE Trans.. On Sig. Proc*, vol. 40 No.3 March 1992, pp. 497-510

[2] G. J. Sandell, "SHARC Timbe Database". Available on the World Wide Web at *http://sparky.parmly.luc.edu/sharc/*

[3] J. O. Smith, "Nonlinear Commuted Synthesis of Bowed Strings," in 1997 Computer Music, Greece.

[4] Several commercial sampling/synthesis products. As an example, see Native Instruments' "Transformator".

[5] N.H. Fletcher, T.D. Rossing, "The Physics of Musical Instruments," Springer-Verlag, 1991.

[6] McGill University Master Samples *http://improv.music.mcgill.ca/newHome/mums/html/mums.html*