Wind instruments synthesis toolbox for generation of music audio signals with labeled partials

Martín Rocamora[1,2] **, Ernesto López**[1,2] **, Luis Jure**[1]

[1]Estudio de Música Electroacústica – Escuela Universitaria de Música
Av. 18 de Julio 1772 – (598) (2) 403 64 44

[2]Instituto de Ingeniería Eléctrica – Facultad de Ingeniería
Julio Herrera y Reissig 565 – (598) (2) 711 09 74

Universidad de la República, Montevideo, Uruguay

rocamora{@eumus.edu.uy,@fing.edu.uy}, elopez@fing.edu.uy, lj@eumus.edu.uy

Abstract. *In this work a methodology is proposed and a set of software tools is released for the automatic generation of synthesized audio files accompanied with labels that describe the temporal evolution of the amplitude and frequency of each one of the partials present. The approach is to synthesize wind instruments sounds using a simple yet effective additive synthesis model based on [Horner and Ayers, 1998]. Some improvements over the original model are implemented and others suggested for future work. In the context of automatic extraction of musical content from audio, this data can be used as ground truth labels for the development and evaluation of algorithms intended for example to estimate partials parameters or track its evolution. This seems an interesting contribution, since manual annotation is a very time consuming task in this situation and a resource of this kind is not available for researchers at present.*

1. Introduction

When developing an algorithm for the automatic extraction of musical content from audio recordings (known as Music Information Retrieval, MIR) it is important to have labeled examples in order to train machine learning schemes, adjust parameters or perform systematic evaluations comparing the results with the annotations. For the learning to be effective and for the evaluations to be representative it is desirable to have a large number of annotated examples at disposal. Depending on the MIR problem considered, the labeled data needed may be different. For instance, the multiple fundamental frequency problem requires the fundamental frequency value of each audio source at a certain time interval.

Usually labels are produced from music recordings by experts that manually annotate the information of interest. Unfortunately, this manual process is in most cases an arduous and time consuming task. For this reason, available resources are quite limited. Among them, the Popular Music Royalty Free Music Database [Goto, 2006] and the data provided by the Music Information Retrieval Evaluation eXchange, MIREX [Downie, 2006], are examples of the most commonly used. Additionally, ambiguities may arise during manual annotation, in which subjective judgements and personal assessment take part [Yeh et al., 2007]. Recently semi-automatic means of labeled data generation were devised, in order to avoid the disadvantages of manual labeling.

An option is to record the live performance of a musical piece whose symbolic representation is available, for instance as a MIDI file. Ideally the MIDI sequence would be an appropriate description of the musical content of the recording. However, in practice differences between performance and symbolic representation do exist, for example due to variations in duration and time onset of notes. Thus, a precise time alignment is needed to adjust the labeling to the performance. This complex problem can be tackled by means of techniques such as dynamic programming [Ellis, 2008]. In addition, other kinds of differences can also exist (e.g. substitutions, deletions or additions of notes), that may require more elaborated solutions.

Another option to perform automatic labeling is to synthesize music from MIDI sequences. There are several ways of generating audio from a MIDI file. MIDI sequencers can be used with sound modules, synthesis software or audio samples of musical instruments. The main drawback of this method is the lack of realism or naturalness compared to real recordings. It is important to notice that MIDI information is sometimes inadequate for labeling. For example, the MIDI file does not describe the temporal evolution of the fundamental frequency during a note. Moreover, recorded samples may not be played in tune. In this case an audio file is usually created for each single track and the labels are obtained by means of monophonic fundamental frequency estimation techniques, as described in [Yeh et al., 2007].

Sometimes an MIR algorithm tries to follow the course of each partial of the sounds in a given recording, for example to identify and segregate the different sound sources or to detect the multiple fundamental frequencies. Manually generating labels for this purpose is an overwhelming task, and to the best of our knowledge there are no resources that provide such data. In this work a methodology is proposed and a set of software tools is released to automatically produce labels in this situation. The approach is to synthesize wind instruments sounds using a simple yet effective additive synthesis model based on [Horner and Ayers, 1998] that generates a dynamic spectrum. Although the synthesis model is not able to reproduce many of the nuances and particularities of wind instruments (e.g some articulations, noisy attacks, blowing noise and legato notes), it generates a wide range of timbres that are clearly identifiable and enables to precisely track the temporal evolution of the amplitude and frequency of each partial. In this way a polyphonic labeled database can be generated by synthesizing MIDI files with this toolbox, which is the main contribution of this work. Some improvements over the original synthesis model are implemented and others suggested for future work.

2. Synthesis of musical instruments

2.1. Timbre

In order to effectively synthesize the sound of a given musical instrument we should be able to recreate its *timbre*. This broad characteristic carries information about the source (such as material, shape, size), type of excitation, etc. Timbre perception is therefore a complex phenomena related to several physical properties. Classical theory of timbre [Helmholtz, 1954] considered the main features to be the waveform amplitude envelope and the spectral magnitude envelope. The former reveals characteristics about the oscillation type (e.g. dumped, forced) and the kind of driving force (e.g. impulsive, continuous). The latter describes how the energy is distributed along the frequency domain. The acoustic system of most musical instruments consists of an excitation source (e.g. vocal folds) and a resonator (e.g. vocal tract). Resonant frequencies, which depend on the size, shape and material of the resonator, emphasize certain spectral regions thus amplifying some harmonics. Therefore the spectral envelope of an instrument typically exhibits characteristic peaks, named *formants*, located at certain frequency regions. Sounds produced by the same instrument (similar timbre) at different pitches, have distinct amplitude relations among its frequency components (see figure 1).

Modern studies of timbre [Risset and Wessel, 1982] showed that during the course of a sound its spectrum changes dramatically (see figure 2). Amplitude and frequency of each partial varies with time and this behavior plays an important role in timbre perception. For this reason, a sound synthesis algorithm must be able to produce a dynamic spectrum.

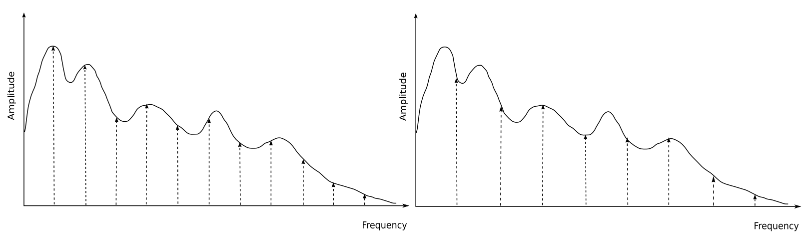

Figure 1: Diagram of the spectrum of harmonic sounds of similar timbre and different fundamental frequency (F0). Transfer function of the resonator exhibits peaks at certain regions. At different F0 the relative amplitude of frequency components is different.

2.2. Additive synthesis by analysis

A traditional sound synthesis technique, named *additive synthesis*, consists in the superposition of sinusoidal components whose frequency and amplitude typically vary with time producing a dynamic spectrum. In order to synthesize a given musical instrument sound, amplitude and frequency of a set of sound oscillators can be controlled with the information obtained by the

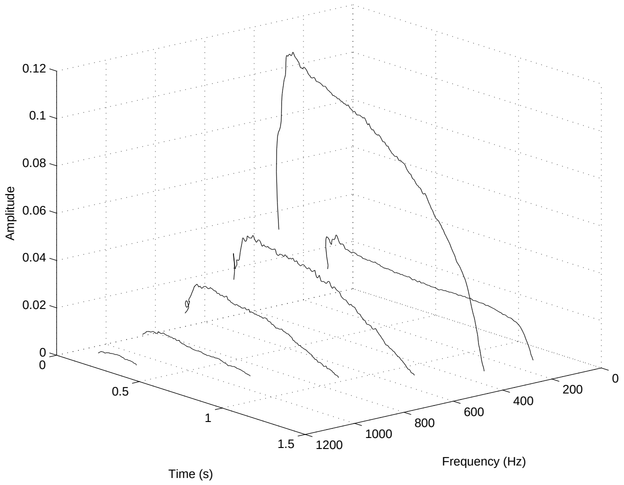

Figure 2: Temporal evolution of the first 6 partials of the sound of a French horn note (G3). Note that the amplitude evolution of each partial is independent. The analysis was performed using SMS software [Serra and Smith, 1990].

analysis of a real sound (like the one shown in figure 2). In this respect, the precise evolution of frequency and amplitude of partials is less important than the global or average behavior [Moore, 1990]. In particular, it is possible to build synthetic sounds that are perceived to be virtually identical to the original recordings, approximating the evolution of partial parameters with line segments or piecewise curves [Dodge and Jerse, 1997] (see figure 3).

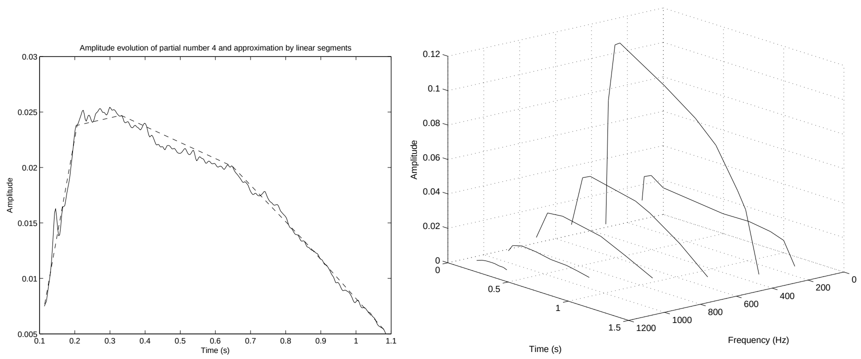

Figure 3: Partials amplitude evolution of the sound of figure 2 approximated by linear segments.

It is important to note that the analysis data is valid only within a small range of frequency and amplitude. This implies that the analysis of a certain note is not generally effective for synthesizing other pitches or dynamics of the same instrument. Changing the fundamental frequency of the note in the synthesis may probably not evoke the desired timbre. This is because in real instruments the location of formants remains the same, so relative amplitudes given by the analysis are changed for a different pitch (figure 1). The relative amplitudes of sound components also varies for different dynamics. Louder notes tend to increase the relative amplitudes of higher partials, making the slope of the spectral envelope less pronounced (see figure 4).

3. Wind instruments sound synthesis

3.1. Wind instruments sound characteristics

Beyond their evident singularities and differences, wind instruments have some common characteristics that motivate the idea of a general synthesis model as proposed in [Horner and Ayers, 1998]. Firstly, all of them consist of a tube resonator and some kind of acoustical excitation that makes the air within the tube to vibrate and resonate at a certain fundamental frequency. The excitation can be generated by the vibration of the performer's lips as in brass instruments (e.g. horn, trumpet, tuba, trombone), or by the vibration of a single or double reed as in woodwind instruments (e.g. clarinet, oboe, fagot, saxophone). In the case of flute type instruments (e.g. transverse flutes, recorders, organ flue pipes) the excitation comes from the effect

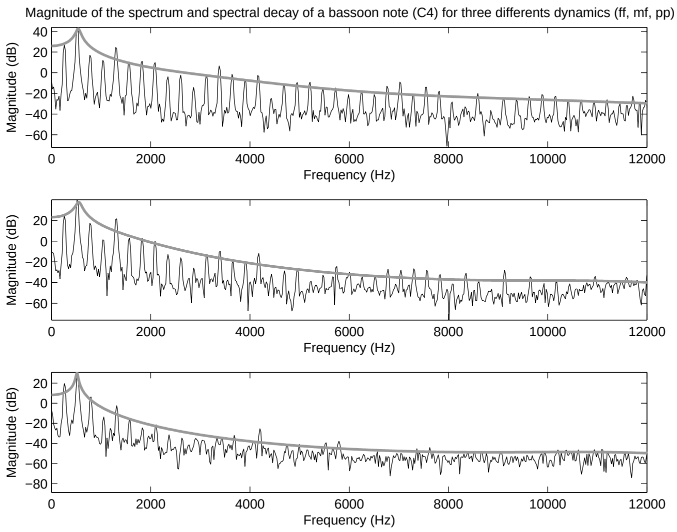

Figure 4: Magnitude of the spectrum and spectral decay of a bassoon note (C4) for three different dynamics (*ff*, *mf*, *pp*). The slope of the spectral decay gets steeper as intensity decreases.

of an air jet striking a sharp edge (edgetone). In either case, the vibration of the air column produces a harmonic spectrum whose components tend to decrease in amplitude with frequency. This spectrum is emphasized in some spectral regions given by the shape of the instrument (e.g. the bell). In this way, wind instruments exhibit characteristic formants [Hall, 2001]. See figure 5 for an example of the spectrum of oboe and clarinet.

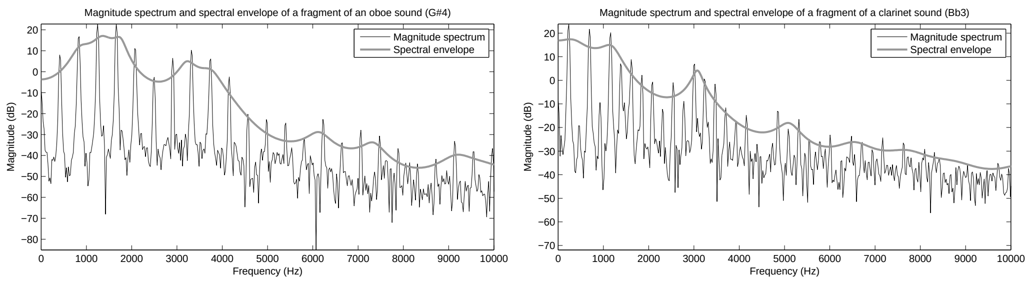

Figure 5: Spectrum of the sound of an oboe note (G#4) and a clarinet note (Bb3). Formants of the oboe give its *nasal* characteristic sound. Regarding the clarinet, its spectrum also exhibits formants and it is noticeable the absence (or relative weakness) of the second and fourth harmonic, typical of this instrument in the lower register [Fletcher and Rossing, 2008].

With regards to the waveform amplitude evolution, it can be usually divided into attack, steady-state and release. During attack and release the envelope can be roughly approximated with an exponential curve (see figure 6). However, some notes may have a more complex behavior, for example a pronounced attack followed by a short decay, steady-state and release.

Wind instruments also exhibit a broadly similar spectral behavior during the course of a note. Considering in turn each of the harmonics from the fundamental frequency, the attack tends to be slower and the release faster (see figure 6), so the harmonics seem to appear one after the other and fade out in the opposite way. Therefore, the sound gradually becomes brighter during the attack, until it reaches its maximum in the steady-state and it gets darker during the release. Additionally, brightness also changes with dynamics in wind instruments as described previously, that is, the relative amplitudes of higher harmonics increase with intensity [Fletcher and Rossing, 2008].

3.2. Contiguous partials grouping synthesis

The synthesis model adopted in this work is the one proposed in [Horner and Ayers, 2000] for the synthesis of the French horn and in [Horner and Ayers, 1998][Horner and Ayers, 2002] for the synthesis of wind instruments, using the *Csound* synthesis language. The idea is to apply an additive synthesis by analysis model with some simplifications. The model is described in the following for the sake of clarity. Firstly, contiguous partials are grouped in disjoint sets in order

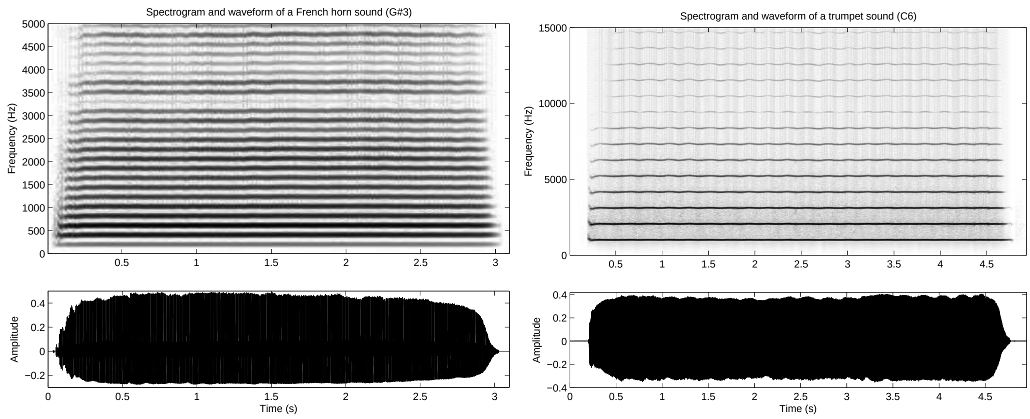

Figure 6: Spectrogram and waveform of the sound of a French horn note (G#3) and a trumpet note (C6). It can be noticed how the harmonics gradually appear one after the other during the attack and fade out in the opposite way during the realease. The waveform amplitude envelope can be approximated with an exponential curve during the attack and release.

to control their temporal amplitude evolution, following a perceptually motivated rule approximately corresponding to the division of the frequency range by critical bands. The first group has only the fundamental frequency component, the second group the second and third partials, the third group from fourth to seventh component and the fourth group has all the remaining harmonics. In this way, the amplitude evolution of each partial, rather than being independent, is such that components of a same group evolve in the same way. Another simplification consists in selecting a single representative spectrum of the steady-state. This spectrum is dynamically modified by changing the amplitude of each group in such a way to produce the behavior of the wind instruments described previously. The selection of the representative spectrum can be performed by means of optimization techniques, but in [Horner and Ayers, 1998] authors point out that picking a spectrum with average brightness gives a similar solution and requires much less computation. Both simplifications significantly reduce the amount of information that is retained from the analysis, by introducing an approximate and less precise model of the time evolution of the partials.

The block diagram of the synthesis model proposed in [Horner and Ayers, 1998] is depicted in Figure 7. The representative spectrum is divided into disjoint sets of partials and the waveform of each group is synthesized. Temporal amplitude evolution of groups of partials are controlled by a series of curves. For the first partial a curve $amp1$ consisting of linear segments is used. The other amplitude envelopes are exponentially related to the first one ($amp2 = amp1^2$, $amp3 = amp1^3$, $amp4 = amp1^4$). This relations assures that higher harmonics attack more slowly and decay earlier, as it is desired for wind instruments. Figure 8 shows an example of a six linear segments envelope for a neutral articulation. The authors propose other envelopes for different types of articulations, such as crescendo, decrescendo and forte-piano. The model also offers the possibility to determine the attack and decay times, which allows a better control of the articulation. The synthesized trumpet note of figure 8 shows the described spectral behavior. In addition, although the original amplitude envelopes are formed by linear segments, the exponential relation of the curves produces a global waveform amplitude envelope that is not linear but approximately exponential during attack and release.

Spectral differences along the instrument register are handled by using in many cases as few as two notes per octave as spectral reference. For a given pitch, the closest spectrum is synthesized, using the amplitude relations among its partials and the amplitude envelopes described above, but with the corresponding fundamental frequency. This approach has the drawback that it modifies the location of formants for pitches that do not match the reference spectrum notes and produces audible discrete timbre jumps along the instrument register.

The synthesis model also includes a dynamic vibrato that makes the instrument sound more natural, as performers typically modify their vibrato during the course of a note. Additionally, the synchronized frequency fluctuation of partials contribute to their perceptual fusion into a single tone [Chowning, 1999]. Vibrato rate changes from approximately 3 to 5 Hz, and a certain random perturbation is added to the starting and ending value so the vibrato is slightly different for each note. Also vibrato depth changes along the note as it is shown in figure 9, and its maximum value is controlled from the synthesis score program.

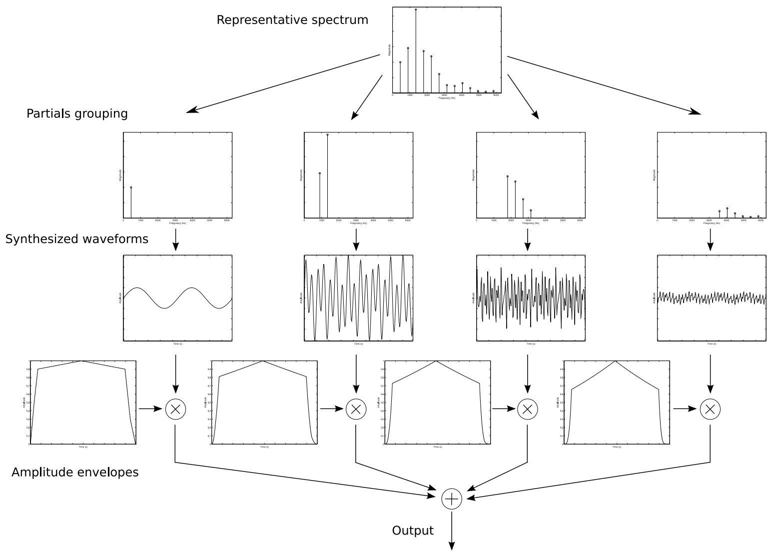

Figure 7: Block diagram of the synthesis system proposed in [Horner and Ayers, 1998].

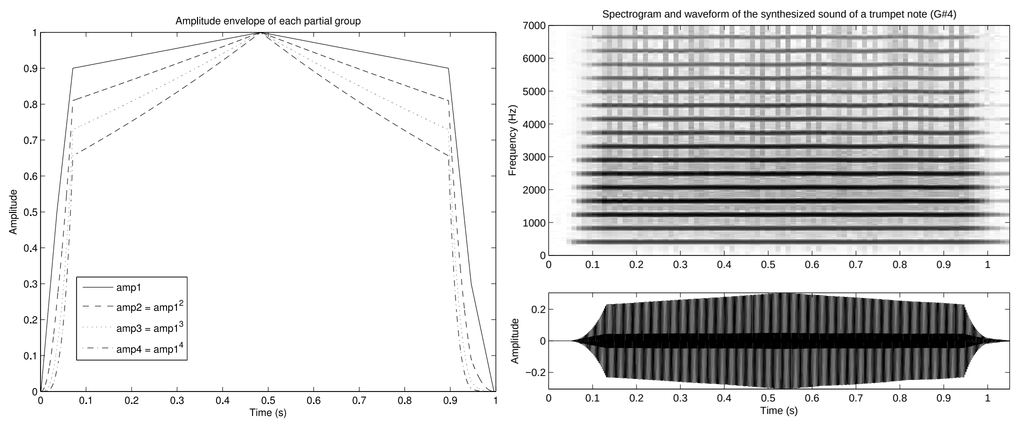

Figure 8: Curves used as amplitude envelopes for each group of partials and synthesized trumpet note. Successive harmonics attack more slowly and decay earlier. The resulting waveform amplitude envelope has an approximately exponential behavior during the attack and release.

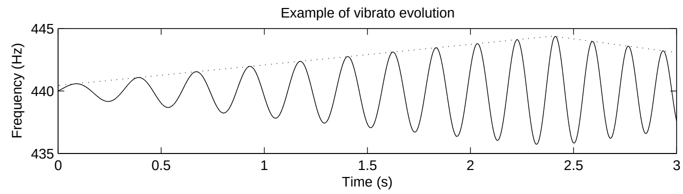

Figure 9: Vibrato evolution along the course of a note. Both vibrato rate and depth are variable.

3.3. Improvements over the original synthesis model

When aurally evaluating the output of the model, the synthetic nature of the sounds produced is recognizable. This is mainly due to lack of natural articulations and discrete timbre variations along the register. An additional deficiency of the model is that timbre changes produced by different dynamics are not taken into account. Therefore, increasing the number of spectral reference notes, including also different dynamics, would seem to offer important improvement. This could be achieved by analysing collections of musical instrument sound samples, such as McGill University Master Samples (MUMS)[1], Iowa Musical Instrument Samples (MIS)[2] and RWC Musical Instrument Sound Database[3]. Ideally, one reference spectrum should be available for each

[1] `http://www.music.mcgill.ca/resources/mums/html/MUMS_audio.html`
[2] `http://theremin.music.uiowa.edu/MIS.html`
[3] `http://staff.aist.go.jp/m.goto/RWC-MDB/rwc-mdb-i.html`

note and dynamic. This issue was explored in two ways in the present work. Firstly by using the SHARC Timbre Database[4] [Sandell, 1991], a collection of steady-state spectrum estimates obtained from the MUMS. Unfortunately, notes provided in MUMS cover only a single dynamic performance. For this reason, an automatic steady-state spectrum estimation procedure was implemented to analyse the MIS database, that includes notes played *ff*, *mf* and *pp*.

A straightforward way to assess the perceptual impact of the synthesis performed with a different reference spectrum for each note is to use the SHARC database. For a given note, it provides the amplitude and phase of the harmonics, corresponding to the sustain or steady-state portion of the tone. Using the SHARC data, several music fragments and chromatic scales were synthesized comprising various wind instruments. Aural tests showed smooth timbre variations of an instrument along its register, substantially improving the results of the synthesis performed with the original model data. During this process some unpleasant notes were identified, that could arise from spectrum estimation errors or erratic labeling and tuning problems of the MUMS database [Eerola and Ferrer, 2008].

In order to include different dynamics into the synthesis model, a steady-state spectrum database was built from the automatic analysis of MIS samples, following a procedure similar to the one applied in SHARC. It consists in selecting a representative spectrum of the sustain portion of the note and estimating the amplitude of the harmonics at this point. This process is depicted in figure 10. The steady-state portion of the note is considered as the longest time interval where the signal energy is greater than 75% of its maximum value. Then the spectrum of each steady-state signal frame is computed and they are summed up in an average spectrum. A representative time instant of the steady-state is determined from the frame whose spectrum most closely resembles the average spectrum, in a least-squares sense. Fundamental frequency is estimated at this time instant based on the autocorrelation function. The actual fundamental frequency may be different from the frequency of the note, due to tuning inaccuracies. After that, the Discrete Fourier Transform (DFT) of a four period length hann-windowed signal frame is calculated. Assuming that the signal is perfectly harmonic and stationary within the frame, partial amplitudes are picked from every fourth bin of the DFT up to 10 kHz.

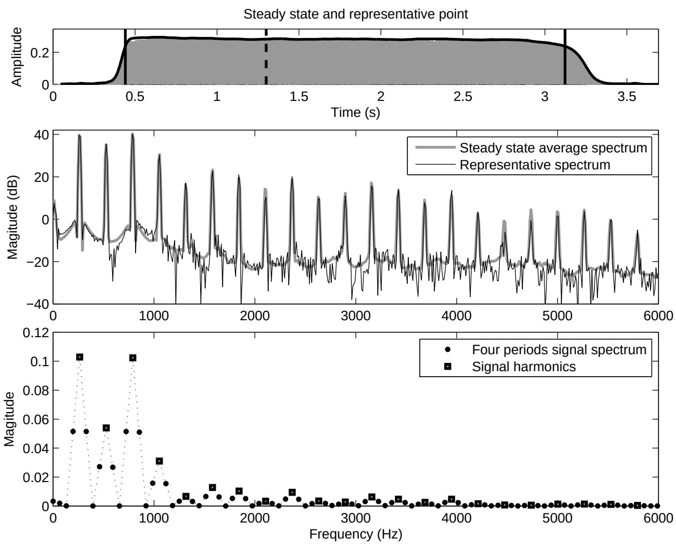

Figure 10: Estimation of a representative spectrum.

When applying this procedure in practice several difficulties were encountered. The criteria used to select the steady-state portion of the note fails for example if the energy exhibits a highly pronounced peak. This may happen in a note with a strong attack, as shown in figure 11. Some simple rules were added to the estimation process trying to avoid this kind of errors. In other cases it is even difficult to establish if a steady-state exists for a certain note. It is reasonable to suppose that some of the bad notes spotted when using SHARC data may come from these problems.

The database built in this form provides the smooth timbre variation previously described for the SHARC data (see figure 12), as well as simulates the natural variation of timbre for

[4] `http://www.timbre.ws/sharc/`

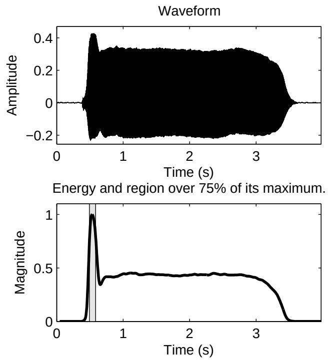

Figure 11: Examples of problems encountered during the spectrum estimation.

different dynamics (see figure 13). Due to the problems related to estimating a single representative spectrum for the whole note, the database inevitably contains some unsatisfactory notes. For this reason, it is necessary to aurally evaluate the synthesis for each instrument trying to identify these note in order to correct them.

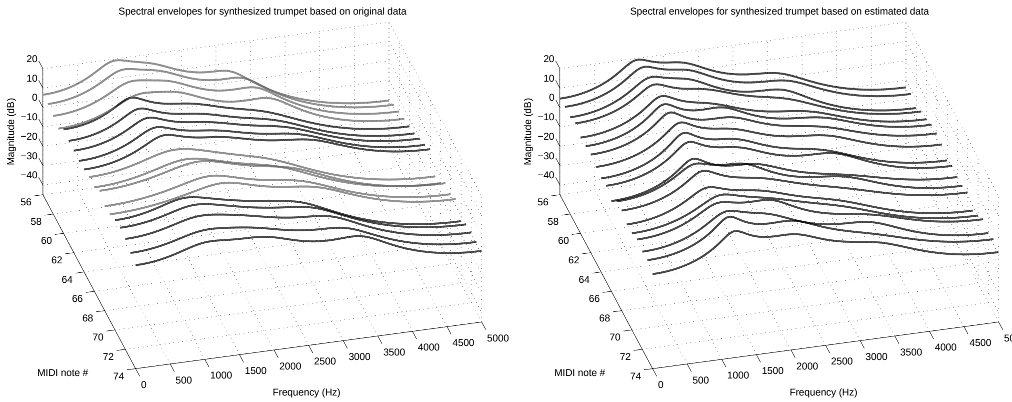

Figure 12: Spectral envelopes for the trumpet synthesized using the original and the estimated spectral data. In the first case, groups of contiguous notes are clearly noticed. Due to this approximation, location of formants is changed from one note to the other. The estimated spectral data presents a much more regular behavior of the spectral envelopes.

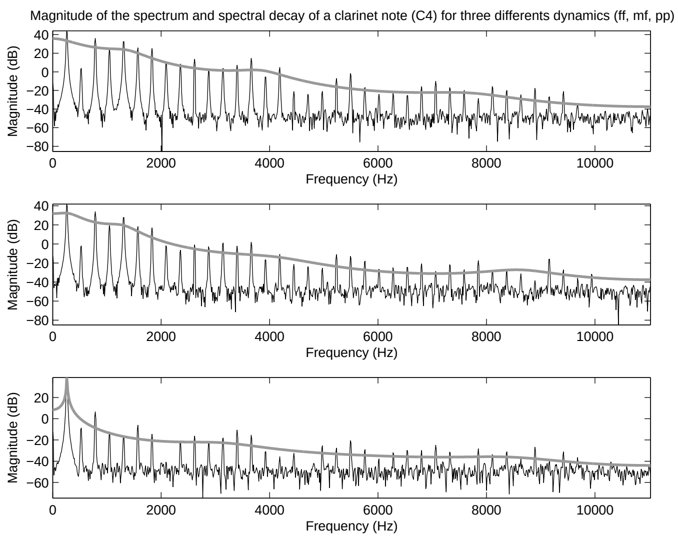

Figure 13: Magnitude of the spectrum and spectral decay of a synthesized clarinet note (C4) for three different dynamics (*ff*, *mf*, *pp*). The spectral decay shows a similar behavior to that of real wind instrument sounds.

3.4. Implementation and examples

The synthesis model was implemented as a publicly available[5] toolbox library in *Matlab / GNU Octave*, based on the *Csound* code in [Horner and Ayers, 2002]. The analysis data used can be chosen from the original data, the SHARC database and the data estimated from MIS audio samples. Figure 14 shows a dependency graph of the functions in the library.

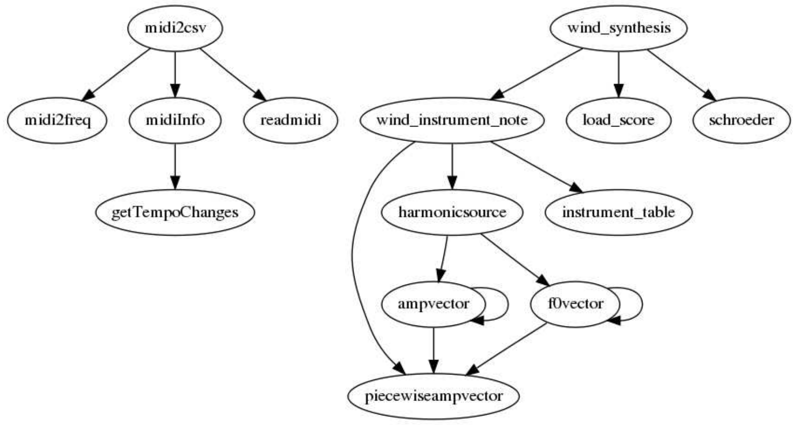

Figure 14: Dependency graph of the functions in the library.

The code in table 1 is an example of usage. The main function is `wind_synthesis`, which returns a vector (`d`) and an array of structures (`note`). The former corresponds to the synthesized audio signal samples, that are also saved to an audio file. The latter represents the data of each note: the name of the instrument, the value of frequency and amplitude of each partial at sample time instants given in a vector of temporal indexes. The function has a set of input arguments that take default values if they are not specified. The first argument is sampling frequency (`fs`), followed by the name of the score input file (`in`) and the name of the output audio file (`out`). It can also be specified if reverberation is added to the synthesis (`reverb`), and if so, also the value of reverberation time T60 (`delay`) and the percentage of reverberated signal added to the output (`pr`). The reverberation uses an implementation of a Schroeder reverberator (4 comb filters followed by 2 all-pass filters) available at [Väänänen, 2000]. Finally an argument controls output verbosity (`verbose`) and another one sets the spectrum data used in the synthesis (`tables`).

```
%% wind synthesis example program
% global parameters
fs = 44100;
verbose = 0;
% reverb paremeters
reverb = 1;    % whether to use reverb or not
delay  = 1.2;  % delay time
pr     = 0.25; % percent of reverberated signal
% output filename prefix
out = 'wind_synthesis-test';
% synthesis tables to use (0: Horner and Ayers - default, 1: MUMS Sharc, 2: MIS Iowa)
tables = 0;
%% score file
in ='scores/Beethoven_Op18-5-var5_quinteto.csv';
% 2 piccolo flutes, 2 oboes, 2 clarinets, 2 bassoons
%% synthesis
[d notes] = wind_synthesis(fs, in, out, reverb, delay, pr, verbose, tables);
```

Table 1: Example code of the library usage.

The score is loaded from a comma separated values file as shown in table 2. If the first row starts with a 0, the second parameter is used to set the beats per minute (bpm, default to 60). Each of the remainder rows correspond to a note. The first parameter sets the instrument number, from the ten wind instruments available (horn, clarinet, oboe, bassoon, flute, piccolo, sax, trumpet, tuba and trombone). Then onset time and duration are specified in beats (or seconds if no bpm is set). The following parameters are the amplitude (a value between 0 and full scale, $2^{16}/2$), the frequency in hertz and the vibrato depth (a value between 0 and 1). Finally the attack and decay time are specified in seconds. It is also possible to perform the synthesis from a MIDI

[5] `http://iie.fing.edu.uy/~rocamora/wind_synthesis/doc/`. Some audio examples are also available.

file. For this purpose a set of functions are provided (adapted from [Schutte, 2008]) to convert a MIDI file into the appropriate comma separated values file.

INSTRUMENT	START	DURATION	AMPLITUDE	FREQUENCY	VIBRATO	ATTACK	DECAY
0,	118,	0,	0,	0,	0,	0,	0
2,	1.500,	0.650,	6000,	294.300,	0.000,	0.050,	0.100
2,	1.950,	0.700,	6500,	490.500,	0.000,	0.050,	0.100
2,	2.450,	0.750,	7000,	392.400,	0.000,	0.050,	0.100
1,	3.100,	10.000,	4000,	294.300,	0.000,	0.060,	0.250
1,	3.000,	10.000,	4000,	196.200,	0.000,	0.060,	0.250
2,	3.100,	2.150,	7500,	588.600,	0.000,	0.050,	0.100
2,	5.150,	0.200,	7000,	654.000,	0.000,	0.030,	0.030

Table 2: Score fragment for clarinet and French horn. First row sets the bpm value.

The library also includes score examples that consider each instrument individually and several polyphonies. Some of them were adapted from [Horner and Ayers, 2002], others come from MIDI files available at *Mutopia* (`http://www.MutopiaProject.org`) under Creative Commons licence and the remaining were prepared for the library from different sources. Figure 15 shows the graphical output of the toolbox for the synthesis of a French horn phrase and a fragment of a piece for oboe, clarinet and bassoon.

4. Conclusions and future work

This work suggests a methodology for creating music audio examples accompanied by labels that describe the evolution of amplitude and frequency of each one of the partials of the sounds present. The method is based on the synthesis of wind instrument sounds following an additive synthesis model described in [Horner and Ayers, 1998]. Despite the aforementioned limitations, the synthesis model is effective since it is possible to clearly identify the wind instrument being synthesized and it also capable of generating a wide range of timbres. Besides, it enables to precisely track the temporal evolution of the partials at low computational cost. Noticeable improvements over the original model, regarding the naturalness of the synthesis, were obtained by automatically analysing the MIS musical instrument sound samples collection in order to gather spectral information of each note played at different dynamics for several wind instruments. In addition a set of software tools is released, that given its ability to synthesize MIDI files, is an useful tool to automatically build a labeled audio database with annotated amplitude and frequency evolution of each partial. This seems and interesting contribution, since manually generating labels for this purpose is an overwhelming task and to the best of our knowledge no such data is available. The methodology of using group additive synthesis for the automatic generation of labeled partials can be extended to other families of timbre [Lee and Horner, 1999], what would provide a richer annotated database. We have used the library in our research and it has proven to be a very handy tool in the development of algorithms for certain MIR applications.

Future work will include the improvement of the spectral data available, aiming a complete chromatic scale along the whole register for each instrument with different dynamics for every note. At present, these conditions are not completely fulfilled because some notes and dynamics are missing in the MIS database and also estimation errors can exist. The current implementation only includes the amplitude envelope illustrated in figure 8. Improvement of the synthesis algorithm would include the implementation of the other amplitude envelopes proposed for the original model. The spectral estimation procedure as well as the synthesis model assume perfect harmonicity of the spectrum. Further research involves the study of the degree of inharmonicity of the available sound samples and the assessment of its eventual impact in the synthesis, evaluating the pertinence of its inclusion in the model. Another interesting area for future research is the addition of an effective noise model to mimic the blowing noise that is clearly perceived in some soft music passages or closely miked solos.

Optimal synthesis results require a carefully prepared score with a detailed control of the amplitude and temporal location of each note, as well as the type of envelope used and the attack and decay times. Even though a well sequenced MIDI file will include the appropriate information about time and amplitude (in the form of velocity) of the notes, the rest of the nuances of articulations and phrasing will be missing. A desired goal is to produce an ample corpus of high quality scores for the toolbox illustrating a variety of musical situations. Finally, the improvement of the toolbox usability and flexibility as well as its implementation as a standalone application is being considered.

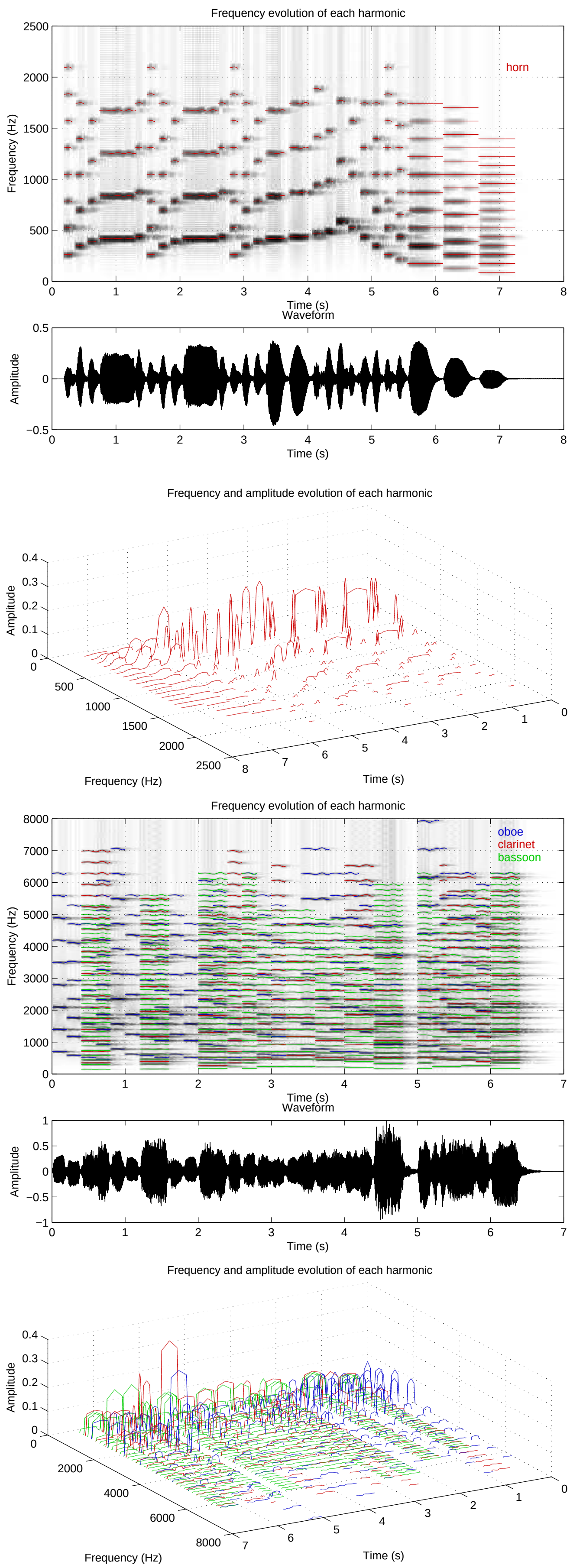

Figure 15: Graphical output of the toolbox for two of the included examples, a French horn phrase and a fragment of a piece for oboe, clarinet and bassoon.

References

Chowning, J. (1999). Perceptual Fusion and Auditory Perspective. In Cook, P. C., editor, *Music, Cognition, and Computerized Sound*. The MIT Press.

Dodge, C. and Jerse, T. A. (1997). *Computer Music: Synthesis, Composition, and Performance*. Schirmer, 2 edition.

Downie, J. S. (2006). *The Music Information Retrieval Evaluation eXchange (MIREX)*. D-Lib Magazine 12 (12). Available from `http://www.dlib.org/dlib/december06/downie/12downie.html`.

Eerola, T. and Ferrer, R. (2008). Instrument Library (MUMS) Revised. *Music Perception*, 25(3):253–255.

Ellis, D. P. W. (2008). "Aligning MIDI scores to music audio", web resource. Available from `http://www.ee.columbia.edu/~dpwe/resources/matlab/alignmidiwav/`. Accessed on 9 August 2009.

Fletcher, N. H. and Rossing, T. D. (2008). *The Physics of Musical Instruments*. Springer, 2nd edition.

Goto, M. (2006). AIST Annotation for the RWC Music Database. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR'06)*, pages 359–360. Available from `http://staff.aist.go.jp/m.goto/RWC-MDB/AIST-Annotation/`.

Hall, D. E. (2001). *Musical Acoustics*. Brooks Cole; 3rd Revised edition.

Helmholtz, H. (1954). *On the Sensations of Tone*. Dover Publications.

Horner, A. and Ayers, L. (1998). Modeling Acoustic Wind Instruments with Contiguous Group Synthesis. *Journal of the Audio Engineering Society*, 46(10):868–879.

Horner, A. and Ayers, L. (2000). Contiguous-Group Wavetable Synthesis of the French Horn in Csound. In Boulanger, R., editor, *The Csound Book: Perspectives in Software Synthesis, Sound Design, Signal Processing and Programming*. The MIT Press.

Horner, A. and Ayers, L. (2002). *Cooking with Csound, Part 1: Woodwind and Brass Recipes*. A-R Editions.

Lee, K. and Horner, A. (1999). Modeling piano tones with group synthesis. *Journal of the Audio Engineering Society*, 47(3):101–111.

Moore, F. R. (1990). *Elements of Computer Music*. Prentice Hall.

Risset, J.-C. and Wessel, D. L. (1982). Exploration of timbre by analysis and synthesis. In Deutsch, D., editor, *The Psychology of Music*, pages 26–58. New York: Academic.

Sandell, G. J. (1991). A Library of Orchestal Instrument Spectra. In *Proceedings of the 1991 International Computer Music Conference*, pages 98–101, San Fransico, California.

Schutte, K. (2008). MIDI for Matlab. Available from `http://www.kenschutte.com/midi`.

Serra, X. and Smith, J. (1990). Spectral modeling synthesis a sound analysis/synthesis based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14:12–24. SMS, Available from `http://www.mtg.upf.edu/technologies/sms`.

Väänänen, R. (2000). Matlab implementation of a schroeder reverberator. Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing. Available from `http://www.acoustics.hut.fi/~riitta/.reverb/schroeder.m`.

Yeh, C., Bogaards, N., and Röbel, A. (2007). Synthesized polyphonic music database with verifiable ground truth for multiple F0 estimation. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR'07)*, pages 393–398. Available from `http://recherche.ircam.fr/equipes/analyse-synthese/cyeh/database.html`.