Physical characterisation of musical instruments

Mathieu Paquier[1], Benjamin Robin[2], Théophane Foggia[3], Jean-Christophe Béra[3]

[1] *Laboratoire d'informatique EA 2215, 20 avenue Le Gorgeu, BP 809, 29285 Brest Cedex*
mathieu.paquier@univ-brest.fr
[2] *INRETS – LTE, 25 av. F. Mitterrand, 69675 Bron Cedex*
[3] *Laboratoire de Mécanique des Fluides et d'Acoustique, CNRS UMR 5509, Ecole Centrale de Lyon*
36 av. Guy de Collongue. BP 163, 69131 Ecully Cedex

Introduction

Automatic recognition of musical instruments is one of problems raised in audio indexing and automatic musical transcription [1]. The prerequisite to any recognition algorithm is the elaboration of relevant features to describe each of the sounds produced by instruments. Current algorithms use features derived from perception studies [2] (attack duration, spectral centroïd, spectrum evolution) and from speech processing (cepstral coefficients, amplitude and frequency modulation, fundamental frequency tracking). These algorithms ([3][4]) provide good results, but feature evolutions are rarely shown, and the used database are often limited. In the present study we recorded our own database with instruments issued from the classical orchestra, traditional instruments from Ireland and Brittany, and different components of drums. Then we observed the evolution of several features over instruments and families.

Experimental set-up

Twenty notes were played on forty-two instruments including those used in classical orchestras, the traditional ones from Brittany and Ireland (keltic harp, diatonic accordion, diatonic clarinet, bombards, tin whistle), and percussions. To compare the recordings of the sounds produced by a same instrument, but played by different people, the musicians involved in this study had various backgrounds from little to solid ones, *i.e.* professional. Moreover, a skilled instrumentist was sometimes asked to play with different models of a same instrument so as to highlight the effects of manufacture and key. Each musician played twenty notes he felt the most representative of his instrument capabilities in terms of pitch, duration, intensity, attack, vibrato,... Giving the instrument players the opportunity to select their own sounds allowed us to get notes out of the usual range of the concerned instrument; Though these notes are frequently heard in the twentieth century music, it is seldom to find them in the available databases. Recordings took place in different rooms to get various reverberation conditions. We used a DPA 4006 microphon and a TASCAM DAP1 DAT recorder.

Features extraction

The following features were calculated from each note played by each instrument:

i) Spectral centroïd: we, first, calculated short-term FFT as a time-frequency representation with a 50-ms sliding window and 50% overlap leading to a 20-Hz frequency resolution. Then, for each time window, each frequency was multiplied by its energy; finally, these weighted frequencies were summed up, and the result was normalised by the total energy in the time window. The spectral centroid for the whole note and the time evolution of spectral centroïd corresponded to the mean of these values across time and to the standard deviation across time, respectively.

ii) Crest factor: this feature was the ratio of maximum energy of the note to the mean one.

iii) Attack duration from waveform: we calculated the duration of attack from 5 to 95% of the max energy of the signal, then from 10 to 90%.

iv) Attack duration and band synchronicity from time-frequency representations: we used two time-frequency representations: the short-term FFT mentioned above, and the outputs of "bark filters" (grouping of consecutive frequency bands from the short-term FFT according to the Bark distribution). The threshold corresponding to the end of the note attack was, first, set at 90 or 95% of the maximum energy of the summed frequency bands. Then, for each band, we calculated the attack duration between the end of the note attack and the start of the attack, at which the band energy reached 5 or 10% of its maximum value. The average of the whole attack durations corresponded to the attack duration of the note waveform shape because of the 25-ms time precision of the time-frequency representations. The standard deviation across the frequency bands corresponded to the band synchronicity : a small standard deviation evidenced synchronous bands during the attack, whereas a high one highlighted their difference.

v) Release duration from waveform: We calculated the time elapsed between the start of the note release and the note end, *i.e.* the difference between the last time the signal reached 95 or 90% of its energy and the one at 5 or10%.

vi) Release duration from time-frequency representations: As previously done for attack duration, we calculated the mean and standard deviation of the release duration across the different frequency bands (grouped in Bark distribution or not). The release duration in a band corresponded to the time elapsed between the global start of the note release (last time at 95 or 90% of the maximum energy of summed frequency bands) and the end of the release in the considered band (at 5 or 10% of the maximum energy in the band).

vii) spectral dispersion: It was expressed by the standard deviation of the energy contained in frequency bands. A low value showed that the bands were of alike energy, whereas a high one evidenced disparity in energy contents.

Results

The different experimental features were analysed with respect to the instrument type and family.

Attack duration from waveform: we could rank the different instruments into several families with respect to the speed of attack from the fastest to the slowest one: i) all the percussions ii) the pianos and pinched strings iii) bowed strings and winds (woodwinds and brass). However, bombards from Brittany (diatonic oboe with a very exponential bell) differentiated from other reed instruments by an attack more alike those of percussions than that of reed ones. The attack durations from the "smoothed waveform" obtained with short-FFTs and Barkbands showed high similarities with the one from waveform. Thus, it allowed no additional separation.

Release duration from waveform: : No distinction between winds, strings and percussions could be based on this criterion; however, within a given family, it permitted differentiation among the various instruments: In the percussion family, all instruments were separable (with respect to the speed of release from the fastest to the slowest one: kick, snare, glockenspiel, toms, cymbals). In the string family, the piano were separated from the pinched strings, more resonant. In the flute family, classical flutes (quite long release) were separated from tin whistles (traditional irish recorders, with a short release).

Band synchronicity from time-frequency representations: short-term FFT and barkbands gave same results: the band synchronicity allowed to isolate the percussion family (with very synchronous bands), and to dissociate some sub-families in instrument families : in the wind family, the double reeds instruments (asynchronous bands) were separated from simple reeds and brass (quite synchronous bands), and in the flute family, the "classical" flutes (very asynchronous) were separated from the tin whistles (quite synchronous).

Spectral Centroïd: No distinction between winds, strings and percussions could be based on this criterion; however, within a given family, it permits differentiation among the various instruments: indeed, among the strings, the pinched strings exhibited a higher spectral centroid than the pianos. Among the flutes, the spectral centroid of tin whistles was slightly higher than that of classical flutes, but very close to that of pinched strings. The spectral centroid height is usually proportional to the pitch of the instrument-played notes. As a consequence, this feature enabled us to distinguish the different components of the drums in the percussion family, and the soprano, alto, tenor and barytone saxophones in the single reed family. On the contrary, with bombards, spectral centroid was ranked in decreasing order as follows: bass Bb bass bombard > Eb tenor bombard> soprano Bb bombard. Moreover, one should note that, among all the studied instruments, bombards displayed the highest centroids. The twenty notes played on a given instrument gave generally very close spectral centroids. This proved the very high reliability of this index, which is very used by our auditory system.

Spectral Centroid evolution (calculated from short-term FFT): For a given instrument we observed great variations of this index from one instrumentist to another one. Thus, it permitted no clear differentiation between the studied families of instruments or instruments.

Crest factor: this feature is high when the sound attack is energy-rich and/or when the sound release is long but energy-poor. So, the very high crest factors of the glockenspiel and cymbals allowed to isolate these instruments from the other ones. Moreover, this feature allowed to separate unsustained instruments (high crest factor : percussion, piano and pinched strings) from sustained instruments (low crest factor : rubbed strings and winds).

Spectral dispersion: this feature allowed to isolate (and to rank from the largest to the smallest spectral dispersion) the flutes, clarinets, bombards, and the snare. The differences between the spectral dispersions of the other instruments were not significant.

Conclusion

Some features allowed to separate instrument families (attack duration, band synchronicity), and other features allowed to separate instruments in a given instrument family (release duration, crest factor, spectral centroid). So a recognition algorithm based on a hierarchical classification could be particulary efficient. Before the algorithm conception, we want to examine other features, like cepstral coefficients, and amplitude and frequency modulations, and to compare feature values obtained from other instrument databases.

References

[1] A. Klapuri, T. Virtanen, A. Eronen, J. Seppänen, Automatic transcription of musical recording, Consistent & Reliable Acoustic Cues Workshop, CRAC-01, Aalborg, Denmark, September 2001.

[2] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, J. Krimphoff, Perceptual scaling of synthetized musical timbres: common dimensions, specificities, and latent subject classes, Psychol. Res. **58** (1995), 177-192.

[3] J.C. Brown, O. Houix, and S. McAdams, Feature dependance in the automatic identification of musical woodwind instruments. J. Acoust. Soc. Am., **109** (2001), 1064-1072

[4] A. Eronen, and A. Klapuri, Musical instrument recognition using cepstral coefficients end temporal features, Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Istanbul, 5-9/6, 2000.