Audio Engineering Society

Convention Paper 7427

Presented at the 124th Convention
2008 May 17–20 Amsterdam, The Netherlands

Piano "Forte Pedal" Analysis and Detection

Antony Schutz[1], Nancy Bertin[2], Dirk T.M Slock[1], Bertrand David[2], Roland Badeau[2]

[1] *EURECOM Institute - Mobile Communication Department, 2229 Route des cretes 06904 Valbonne Sophia-Antipolis, France*

[2] *Telecom ParisTech - TSI Department, 46 rue Barrault - 75634 Paris, France*

Correspondence should be addressed to Antony Schutz (antony.schutz@eurecom.fr)

ABSTRACT

Pressing the sustain pedal is one of the main musical gestures in a piano performance. It influences the sound produced by the instrument, and consequently the efficiency of any automatic system analyzing it. This paper aims at highlighting several features, observable on piano sounds played with pedal, which could be useful in the task of detecting the sustain pedal, and analyzing notes played while it is pressed.
Here, we consider two hypotheses, derived from physical acoustics considerations and signal observations, that could help discriminate between notes played with and without the sustain pedal. First, the sustain pedal is found to increase the decay time of partials. This effects dominates the behavior of the partials, not only in duration but also in terms of spectral evolution. Secondly, when the sustain pedal is used, a noise floor appears, for all notes of the piano. Those two effects are investigated in this work, in particular through a method based on a "harmonic plus noise" decomposition. The possibility to use those features as a base for a detection system is discussed.

1. INTRODUCTION

Music transcription is the process of creating a musical score (i.e. a symbolic representation, such as a MIDI file, of the music within) from an audio recording. In the traditional sense, automatic transcription implies the estimation of several features such as the pitch and duration of individual notes.

But music can not be reduced to a succession of notes, and an accurate transcriptor should be able to detect other performance characteristics, such as slow tempo variations or, in the particular case of the piano, the use of pedals. Usually pianos have two or three pedals [1], among which the most frequently used is the sustain pedal, also called forte pedal.

In a piano, the sound generation mechanism works as follows: when the musician presses a key, a hammer strikes the string (or actually between one and three strings, depending on the key) and this interaction triggers the note. When the key is released, a damper comes to stop the vibration of the strings and the note fades out. When the sustain pedal is pressed, all the dampers of the piano are kept raised; this allows the strings to keep vibrating after the key is released, and allows strings associated to other keys to vibrate, due to sympathetic resonance, and coupling *via* the bridge. If several notes are played with the pedal, they will be mixed with a longer duration. A second effect has yet to be noticed. As a matter of fact, the two higher octaves of the piano do not have any damper, but the use of the pedal still has an influence on the sound. For this range of notes, the note does not last longer with or without the pedal, but a natural reverberation due to the resonance of the sound board appears and this sound leads to an additional floor noise.

Similar observations can be found in previous work. [2] proposes a polyphonic piano transcription system which detects and takes into account the use of the pedal. The detection of the pedal is based on an estimation of the noise floor. It is estimated as the mean value of the Discrete Fourier Transform (DFT) magnitude over the analysis frame, but only on frequency bins considered as "not active" in the frame (not associated with an actually played note - these frequencies are determined by a varying threshold). Another modelling of the sustain pedal can be found in [3, 4]. Through the analysis of middle-range piano notes, played *legato* with and without the pedal, the authors point out three features that should be able to discriminate between notes played with and without the pedal, and be useful for piano synthesis: noise floor, decay time of the partials and amplitude beating.

In the following, we first describe the database that was built up to study the effect of the pedal on piano notes. We then present how features are extracted from this database, leading to observations and results on their discriminative power. We conclude our work with discussing the oppportunity to use them as sustain pedal detection features.

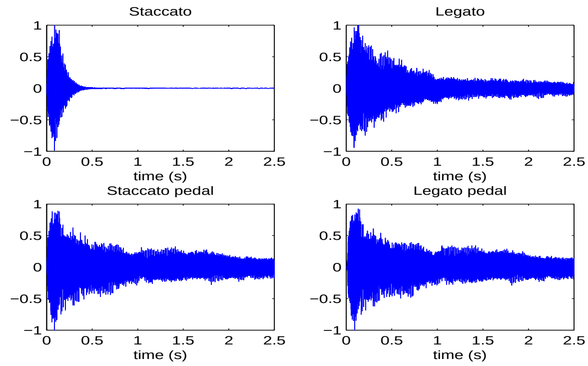

Fig. 1: Examples of waveforms of *staccato*, *legato*, *staccato+ped.* and *legato+ped.*, note D_2

2. DATABASE

Special recording was done in order to study the effect of the sustain pedal. Two identical microphones (omnidirectional electrostatic Shoeps) were placed on the right side of a grand piano (grand piano Yamaha C1) at one meter from the sound board and the sound was digitalized at a sampling rate of $44.1KHz$ and encoded with 16 bits, through an Edirol UA5 soundcard. This configuration was chosen in order to gather a maximum of the resonance generated by the sound board when the pedal is pressed.

We initially considered that the actual gesture of the musician could have an importance, and we decided to distinguish between notes played *staccato* (short strike on the key) and played *legato* (the key is kept pressed on). For the *staccato*, since the strike is short, the damper takes only very little time to go down, whereas we have the opposite for notes played *legato*. The database is thus composed of four categories of notes: *staccato* without pedal (*staccato* in the following), *staccato* with pedal (*staccato+ped.*), *legato* without pedal (*legato*) and with pedal (*legato+ped.*), some corresponding waveforms are illustrated in Figure 1. We recorded single tones from low to high frequency range of the piano, with and without the use of the sustain pedal. The interval between each note is a fourth (C to F for example) and each note was played in the four configurations previously described. It lead to a database of all in all 200 note recordings.

3. FEATURE EXTRACTION

As we want to study two features, one being specifically linked with the sinusoidal part of the note (decay time of the partials) and the other concerning the noise (noise floor power), it seems natural to perform an "harmonic plus noise" decomposition [5] before feature extraction itself.

3.1. Pre-processing

Since for real recordings the background noise is seldom white a preprocessing step is applied. The original spectrum is whitened by means of autoregressive modeling (AR) of the background power spectral density. In order to increase the number of points in the spectrum we use zero padding. The background spectrum is obtained by median filtering and inverted by a Finite Impulse Response (FIR) filtering, at the end this operation is compensated by AR-Filtering. However the purpose is to study piano tones. The very large range of frequencies covered by the piano (*88* notes from *27.5Hz* to *4186Hz*) makes it difficult to have the same efficiency for all the notes. In order to increase the resolution each studied note is slightly decimated according to its range of frequencies. Since we are in a monophonic case we have used a correlation method for this purpose.

3.2. Harmonics amplitudes tracking

For the study of the envelopes of the partials we have used Fast Sequential LS Estimation [6]. This method is an adaptive algorithm used for the estimation of slowly varying amplitudes. It assumes the frequencies are known in advance and gives a continuous evolution of each partial. It takes into account the sinusoidal nature of the data and because it uses a rotational invariance technique, has a low complexity. First of all, the preprocessing procedure explained above is applied to the signal. Then in the whitened magnitude spectrum, we find the frequencies by searching for local maxima (peak picking). Finally we use them as inputs for Least Square Estimation. Figure 2 shows an example of the evolutions of the first three harmonics for a note played with and without the sustain pedal. The Pedal has a dominating effect on the envelopes. During the attack the evolution is the same for the two cases but after $100ms$ the behavior changes. For this note the decay and release times change and a beating appears on the first harmonic.

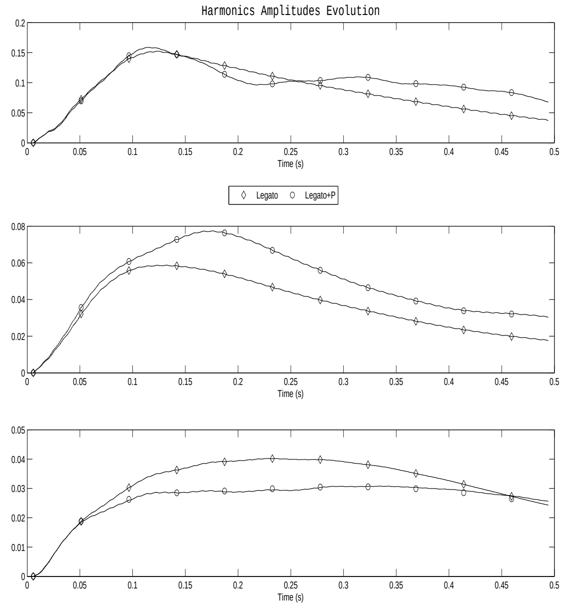

Fig. 2: Evolution of the amplitudes of the first three harmonics

3.3. "Harmonics plus noise" decomposition

The audio signals are modeled as a sum of sinusoids with time varying parameters. However, in this model the nature of the instrument is completely ignored [1]. When we extract the sinusoidal part of the signal, we obtain a noise part composed, in the best case, of the background noise and the instrumental noise. For the piano the instrumental noise is essentially due to the strike of the hammer on the strings and the resonance of the sound board. The decomposition is provided by the Fast LS algorithm.

Figure 3 shows the Short Time Fourier Transform (STFT) of the total signal and of the noise only part in the decomposition for the *legato* and *legato+ped.* cases. As expected the two STFT are similar but not their noises. For the *legato* the residual noise is due to the attack and, except some slightly resonance, fades out very quickly. The residual of the *legato+ped.* is a more sustained one and lasts longer. As for the amplitude evolution during the attack the two cases are undistinguishable below a time interval shorter than $100ms$.

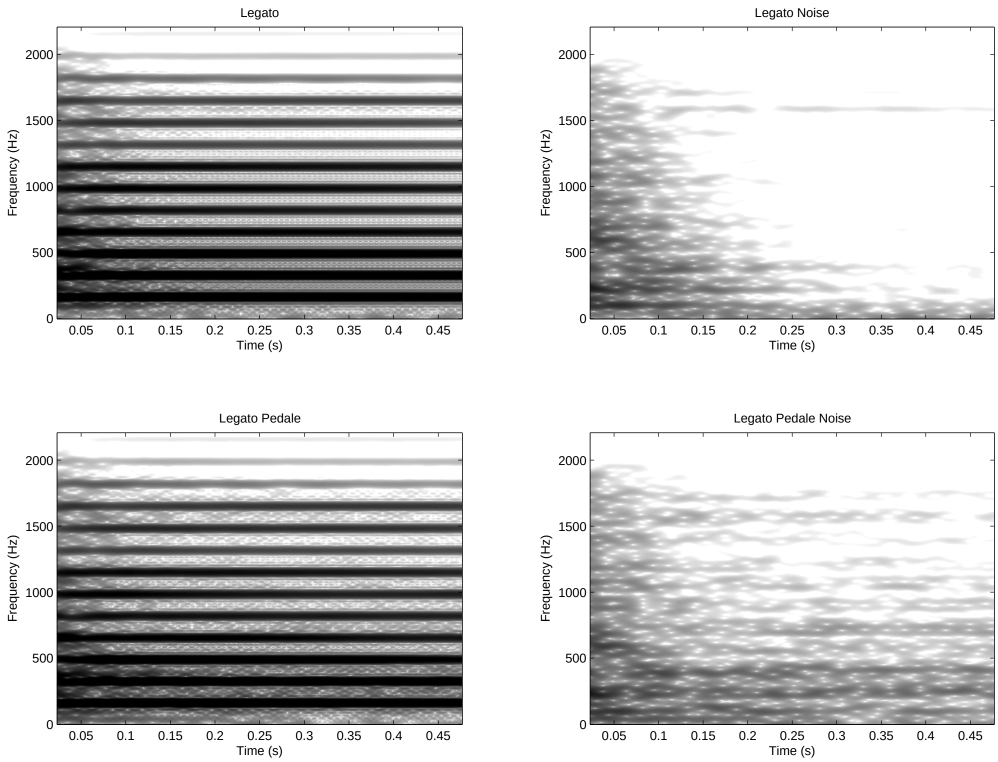

Fig. 3: STFT of total signal (left) and noise only (right) in a Harmonics plus Noise decomposition for the *legato* and *legato+ped.*.

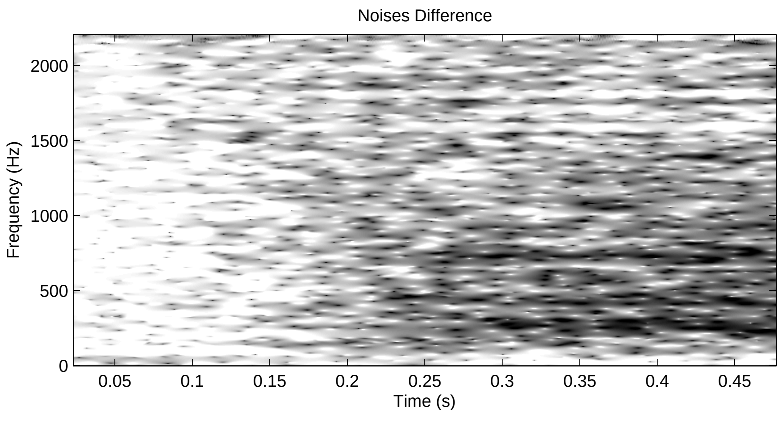

Fig. 4: Difference of the noise spectra for the *legato* and *legato+ped.*.

3.4. Noise estimation

Figure 4 shows the difference of the noise spectra of the *legato* with and without the sustain pedal. Only the noise of the sound board remains, the noise due to the attack is removed due to the differencing.

As it seems impossible to detect the resonance of the pedal before the end of the attack, we start the study at $200ms$ after the begining of the note. The duration studied is also $200ms$, the signal is then normalized in energy. As the presence of the noise of the Pedal is constrained to the low frequencies, we first decimate the signal by a factor 20 then we get the noise by the harmonic plus noise decomposition. We model the noise as an autoregressive process (AR) of the first order [7], by using the Levinson algorithm, to obtain the shape of each noise spectrum. The result is illustrated in Figure 5, which will be discussed below.

4. OBSERVATIONS AND RESULTS

4.1. Decay time of the partials

As stated previously, the use of the Forte Pedal increases the decay time of the partials. Raising the dampers leaves all the strings free to vibrate sympathetically with whichever notes are being played. So all notes played with the pedal have a longer duration. This can lead to some polyphonic playing configurations that would be humanly impossible without use of the Pedal :

- More than ten notes can be present.
- Some configurations are spatially impossible.

The use of this information for detecting the pedal will lead to a complex system that analyses and tracks all notes and then decide if it is possible without the use of the Pedal. Another point which is less important is that this kind of analysis will miss some events, such as a shorter use of the pedal, or will create confusion with longer notes. This is less important because the Forte Pedal is generally used for creating some *Legato* event, not easily realisable in practice.

4.2. Noise floor

Figure 5 shows the result of the AR modelling of noise obtained as in Section 3.4. White lines separate the Pedal cases from the non Pedal cases. It appears that the AR has a flatter shape for the pedal with a slightly lower power in the low frequency range. The bottom of Figure 5 shows the total power in each AR spectrum. Using 30 measurement data we have trained a threshold that separates the two cases. The power of each AR of the training data was computed and the result shows a separation between the two cases. We applied the same threshold to the other data and put the results on the same figure with a point for the notes estimated to be played with the sustain pedal. Results :

- 3 out of 85 pedal noises are interpreted as non pedal, around 96.5 percent.
- 21 out of 85 non pedal noises are interpreted as pedal, around 75 percent.

So a total error rate of 15 percent is obtained. In spite of the simplicity of the method used we achieve a good detection rate. Note that the results may be highly dependent on the harmonic plus noise decomposition used.

5. CONCLUSION

In this paper, we have presented an analysis of the Sustain Pedal piano effect in the monophonic case. We have shown some features, by using recent signal analysis methods, like the envelopes of the partials and the residual noise in a Harmonic plus noise decomposition. When noise is present, the detection of the sustain pedal achieves a good rate but this requires a special recording which is not necessarily used in commercial recordings. The results are encouraging, but still a lot of work is required. Some important future directions include methods based on the evolution of the amplitudes of the partials, the extension to the polyphonic case and the use of a multi class classification method for the detection.

6. ACKNOWLEDGEMENTS

Eurecom Institute's research is partially supported by its industrial members: BMW, Bouygues Télécom, Cisco Systems, France Télécom, Hitachi Europe, SFR, Sharp, STMicroelectronics, Swisscom, Thales. The work reported herein was also supported by the European Commission under contract FP6-027026, Knowledge Space of semantic inference

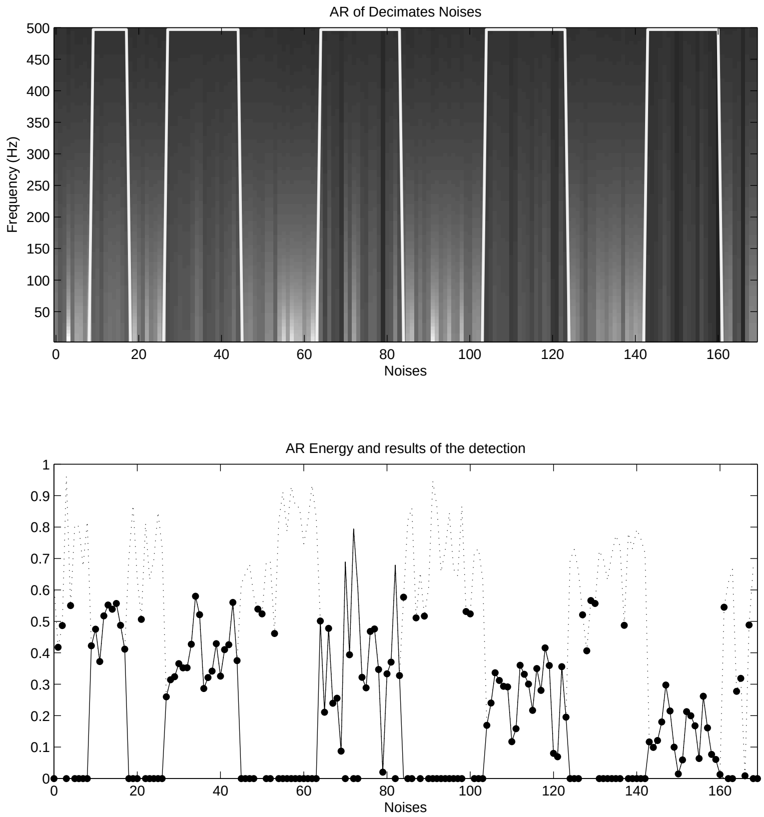

Fig. 5: Top : Autoregressive modeling of the Noise for 170 note recordings. A white line on top indicates notes with Pedal. Bottom : power of the AR model : For notes with Pedal (solid line) and notes without Pedal (dashed line). The dots indicate the notes that are estimated to be notes with Pedal

for automatic annotation and retrieval of multimedia content - K-Space - and by the *Institut Télécom* Crédits Incitatifs GET2007 project TAM-TAM.

7. REFERENCES

[1] Fletcher N.H and Rossing T.D, “The physics of musical instruments”, Second Edition, Springer-Verlag, 1998.

[2] Barbancho I, Barbancho A.M, Jurado A, Tardon L.J, “Transcription of piano recordings,” Applied Acoustics, vol. 65, no. 12, pp. 1261-1287, Dec. 2004.

[3] Lehtonen H.M, Penttinen H, Rauhala J and Valimaki V, “Analysis and modeling of piano sustain-pedal effects” The Journal of the Acoustical Society of America, vol. 122, no. 3, p. 1787, 2007.

[4] Lehtonen H.M, “Analysis and parametric synthesis of the piano sound”, Master thesis, Helsinski University of Technology, Finland, 2005.

[5] Serra X and Smith J, “Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition”, Computer Music J., vol. 14, no. 4, pp. 1224, Winter 1990.

[6] David B and Badeau R, “Fast sequential LS estimation for sinusoidal modeling and decomposition of audio signals”, 2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics WASPAA2007, New Paltz, New York, USA, 21-24 oct. 2007, pp. 211-214.

[7] Jackson, L.B., “Digital Filters and Signal Processing”, Second Edition, Kluwer Academic Publishers, 1989. pp.255-257.