Real-Time Detection and Visualization of Bad Clarinet Sounds

Aggelos Gkiokas, Kostas Perifanos, Stefanos Nikolaidis

Institute for Language and Speech Processing

Introduction

The growing use of computer software for music teaching led research to a new evolving domain, Music Visualization. Music visualization can serve different purposes in the context of music education.

The aim of this work is to provide a real-time music 3D-visualization tool for clarinet sound in the context of music education. The feedback provided in real-time should be short and simple, avoiding to distract the students and helping them to go on despite any errors. Furthermore, the tool must help the student to gain a perception of their progress as the time goes by. Our approach focuses on students from beginning to intermediate level.

Clarinet Bad Sounds

Hollow Tones:

The main cause of a hollow note is the bad airflow in the clarinet. The main attribute of a hollow note is that the energy of the individual harmonics is lower than the normal.

Squeak Tones:

The main cause of a squeak is saliva (into the clarinet or when the reed gets calcified) or when students press and bite the reed resulting no free vibrations. In a squeak note all the partials amplitudes become much higher than the normal.

Unstable Tones:

Unstable notes have many causes such as insufficient amount of airflow for the specific tone or not firm embouchure. Instability can be either pitch instability or RMS energy instability. Both can be easily detected by calculating the standard deviation of pitch and RMS-energy within a note (or part of note)

Bad Sound Detection

➢**Features Used:** Partial amplitudes up to the 6th for each frame divided by the amplitude of the first, denoted by $\{f_j\}, j=1..5$

➢**Training:** In the training phase we fit a Gaussian distribution to each feature from professional recordings. Because of the dependency between pitch and partials we train one model for each pair of feature-tone, denoted by

$$p_{ij} = p_{Gauss}(\mu_{ij}, \sigma^2_{ij}, f_j), i=1..N, j=1..5$$

➢**Error Detection:** For each frame we compute "Q" value as

$$Q=\sum_{j=1}^{5}[\mathrm{sign}(f_j-\mu_{ij})(1-p_{ij})^4]$$

The polynomial power of 4 is used to smooth small variations of $1\text{-}p_{ij}$ around zero. The sign function in each clause has the role to define if the corresponding feature contributes for the sound to be heard more as a squeak or hollow. If Q is positive, means that we probably have a squeak sound, if negative a hollow sound. Closest to zero is Q, the better the sound is.

➢**Other Features:**

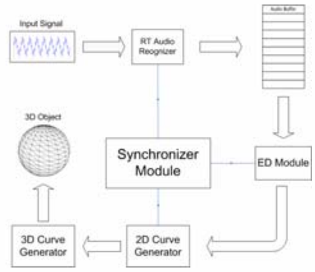
➢**Time-Averaging:** Between two consecutive visual frames we take the average of the "Q" values to judge an acoustic segment.

➢**Onset-Offset Discarding:** Onsets have different statistical properties, thus In a very simple fashion, we discard onset frames, by ignoring the first frames of each note. The same properties stand for the offset of each note. We handled this situation in terms of smoothing: When we have decaying in RMS energy on the signal, implying the note ends; we limit Q from changing value greater than a certain ratio.

➢**Different Level of Students:** To cope with different levels of students we substitute standard deviation for each component by a multiple of it by a value α. The lesser α is, the more sensitive is system to mistakes, allowing the to adjust it.

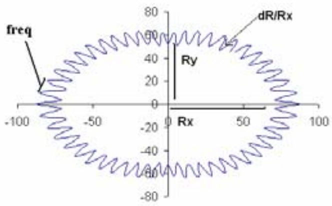
System Overview

Real-Time Audio Recognizer (RTAR) reads streamed data from the microphone as the student performs the musical piece. With a conventional front-end processing scheme RTRA process a window of 25 ms long every 10 ms with a 60% overlap. For every window it processes, RTAR writes output data to the Audio Buffer and sends a message to the synchronizer module that a new frame is processed. The synchronizer activates the Error-Detection (ED) module. ED reads the Audio Buffer and computes the "Quality" of the sound. Every 4 iterations of this procedure the synchronizer sends a message to the 2-Dimensional curve generator which produces the 2D curve. Finally the 2D curve is fed to the 3D-Curve Generator which draws the final shape.



The Visual Model

The main idea is to represent a note as circle. This circle has four attributes to control (plus the color, a total five). These attributes can be shown in figure below. Changing the values according to the student's performance produces a meaningful shape evolving over time.



Visualizing a Squeak Sound

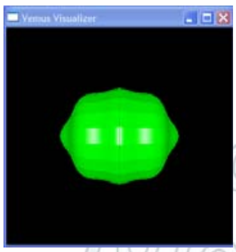
When a frame is classified as a squeak, the shape is drawn as "craggy" or "rough". As more squeak a frame is, the rougher the circle should be. The rules that determinate the circle's attributes values are the following: *Ry/Rx=1* and attribute *freq* is high valued, and increases as squeakness increases. *dR/Rx* is proportional to squeakness and *R* is proportional to RMS energy.

Visualizing a Hollow Sound

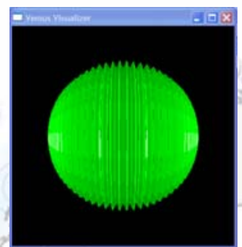
A Hollow note is represented as a more "flabby", "sleazy" shape, as shown in Figure 7. *Ry/Rx* is decreasing as hollowness increases and attribute *freq* is low valued and decreases as hollowness increases. *dR/Rx* is proportional to hollowness and *R* is proportional to RMS energy.

Pitch and RMS Instability

The RMS instability is directly related with the sphere shape, because of the proportional relationship between RMS energy and *Rx*. Therefore an RMS unstable note is directly shown.



3D representation of a hollow tone



3D representation of a squeak tone

This work has been carried out under the VEMUS project which has been partially supported by the European Community under the Information Society Technologies (IST) priority of the 6th Framework Programme for R & D.