Musical instrument recognition and tone identification

Literature survey by: Keerthi C. Nagaraj
Date: 5th March 2003

Objective:

To define a musical instrument in terms of its spectro-temporal characteristics, and use the information to track the notes played by the instrument in a given recording.

Applications:

- Content based audio retrieval
- Automated Music Transcription
- Karaoke track(*Music Minus One*) production,
- Computer participation in live performance, etc.

Background:

The identification of music is done among human beings based on:

- Recognition of the instrument – timbre, pitch, spectral brightness, and other spectro-temporal features
- Recognition of the tone – melody line, bass line, pitch to pitch transition, rhythm

To make a computer do the same thing, we model a musically trained human-like auditory system.

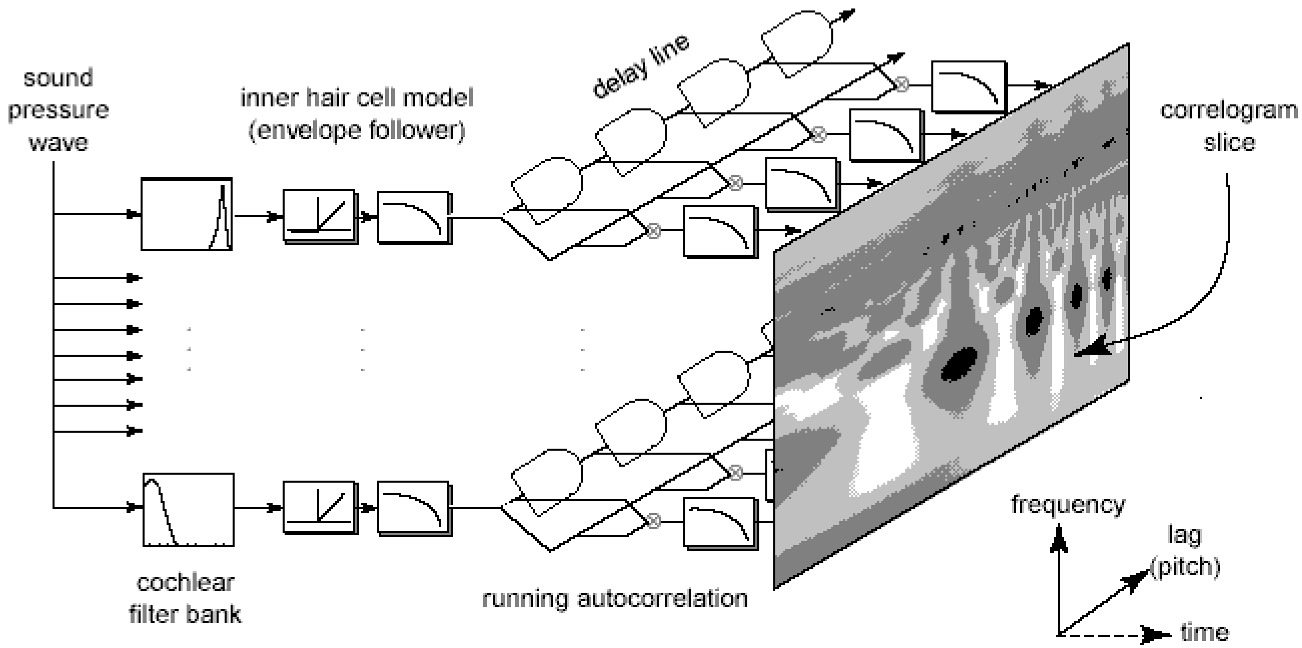
Issues involved:

- Timbre – mathematically ill defined quantity, different results by musicians and non musicians
- Tone– multiple overlapping frequency components, noisy environment etc

Approach to the solution in 3 stages:

- Cochlear Modeling
- Feature extraction
- Pitch detection and tracking

STAGE 1: Modeling the Cochlea



Log lag frequency model of the human ear with autocorrelogram extraction

Meddis, R. and Hewitt, M.J. (1991) 'Virtual pitch and phase-sensitivity studied using a computer model of the auditory periphery: I pitch identification', Journal of the Acoustical Society of America, 89, 2866-2882

Stage 2: Feature extraction

Spectral features: spectral centroid, Average relative spectrum, High frequency roll off rate, intensity, spectral envelope (becomes important later)

Pitch, vibrato & tremolo features: pitch range, absolute strength and relative strength and phase (in comparison to vibrato)

Attack features: Onset asynchrony, inharmonicity, etc

Brown, J.C. (1999). ``Computer identification of musical instruments using pattern recognition with cepstral coefficients as features'' J. Acoust. Soc. Am. {\bf 105}, 1933-1941

Stage 3: Pitch tracking and Note identification

The method proposed in this paper provides a tradeoff against the computational rigor involved in the Meddis and Hewitt model.

The key differences are:

- warped linear prediction for pre-whitening filters,
- Uses critical band frequency resolution rather than uniform frequency resolution
- 2 channels for computational efficiency and almost same performance as that of Meddis and Hewitt model (40-128) channels.

Demerit: It does not use inputs from features extracted!

Another interesting approach:

Goto. M, "A Robust Predominant-F0 Estimation Method For Real-Time Detection Of Melody And Bass Lines In Cd Recordings" ***ICASSP 2000 (2000 IEEE International Conference on Acoustics, Speech, and Signal Processing) Proceedings*****, pp. II-757-760, June 2000.**

Tolonen, T., and Karjalainen, M., "A Computationally Efficient Multipitch Analysis Model," IEEE Transactions on Speech and Audio Processing, vol. 8, no. 6, pp. 708-716, November 2000

What I propose to do:

- A “selective attention mechanism” can be devised to track multiple sources simultaneously using sound source discrimination based on the features extracted.
- Comparative analysis of the pitch tracking mechanisms with the main aim of application to music transcription will be undertaken.
- A working re-synthesis model will be designed to compare the performances of different models.
- Possibility of enhancements in the pitch tracking method with the addition of prior knowledge about the musical instruments being played will be explored.
- Analyze the performance improvement in comparison to the traditional pitch analyzer