A Review of Automatic Description Systems (Dixon) *Naru Sundar, ISE 575*

This paper was an overview of the state of rhythmic description research. It covered a number of different topics and what particular directions were being pursued in each. The first section tackled representations of musical rhythm. It was noted that different levels of rhythmic structure were important to different kinds of people, such as composers, researchers and listeners. The first type of representation was metrical structure, formalized by Lerdahl et.al. in The Generative Theory of Tonal Music. This paper described a formal musical grammar with certain rules for how beats are noted in a score. Mention was also made of various musical psychology studies that looked at how listeners perceive the different levels of musical structure in a piece. Within a metrical structure, tempo, being the rate of beats at a given level of representation is also an aspect of the score which can be perceived. It is not necessarily constant as many forms of music show, and the depth of tempo perception is tied into many factors including how much musical training a listener has. Tempo is also a local phenomenon and does not necessarily have to be constant across the piece. One of the issues of more constrictive forms like that described in GTTM is that many aspects of music involve expression via timing, and timing is a more irregular aspect of most music. Different approaches deal with shifting timing patterns, such as describing tempo as a function of time, or reinterpreting the onsets in terms of the shortest pulse interval. Most musical psychology studies indicate that listeners perceive subtle changes in timing. The paper then proceeded to describe a general form for rhythm description engines which consisted of: feature list creation from audio, pulse induction, pulse tracking and rhythm parsing. Each block was discussed in detail.

Feature list creation was the translation of acoustic signal information into some set of features such as time, duration or amplitude of onsets. Onset time extraction is a very common technique in the literature and a number of cited papers used it. Inter-onset-arrivals or durations of onset notes were another type of feature. Again MIDI input allowed this quite easily, but raw audio proved more difficult. Other onset features such as amplitude and pitch were briefly discussed. Chord finding was another type of feature, constructed easily from MIDI but difficult in raw audio. As we saw in the Goto paper, an FFT based breakdown can be used to extract chord information. One rather unique type of feature was the concept of chunking the audio data into "frames." The author of this particular paper sought to escape overly focusing on specific abstractions such as notes and pitches. From a frame perspective other features become more salient, such as measuring the energy of the chunk or measuring the variation of energy across the chunk. I found the frame concept to be interesting but rather limited in its possibilities, due to the previously ascertained difficult of breaking apart multiple instruments from mixed raw audio.

The process of pulse induction was then looked at, which is the process of identifying periodic structures in the feature lists. Pulse selection is one mechanism for identifying pulses, wherein inter onset arrivals are used to derive pulse schemes for the entire piece, and chosen based on best fit. This process was similar to the Dixon paper from last week. A number of similar techniques based on this were cited. The second technique mentioned was that of computing a periodicity function, of which FFT analysis and autocorrelation were mentioned as specific types. Various variations on autocorrelating functions were given that modified the coefficients returned from the autocorrelation. The Scheirer technique of using tuned filter banks was also discussed. Since periodicities can vary across a piece, mention was made of techniques to quantize the feature list to account for shifts in periods between events throughout the song.

Pulse tracking was discussed as a complementary feature to pulse induction. Pulse tracking seems to involve more real-time algorithms that can track periodicities as the song progresses. Two types of trackers were discussed, sequential trackers that let each note influence the tracker, or predicted beat trackers that only considered the effect on notes that were tolerably predicted. The prediction scheme involved tracking the next note time as the current note time plus the period. One interesting method by Allen and Dannenberg that was discussed involved adding a decay and confidence parameter to this style of pulse prediction, and using hypotheses of pulse trackers that were adaptively selected as more events are processed. This scheme sounded like a real-time version of the Dixon paper.

Lastly the paper looked at further elements of rhythmic structure. This included the problem of time

signature determination. this problem neatly breaks down into two parts, the determining of the length of a beat and the number of beats in a measure. Determining the former is feasible by studying note onset intervals, but it seemed that determining the latter was more difficult since the bar separation was more ad hoc and previous papers have shown how difficult it is to determine these structures.