Evaluation of Performance-to-Score MIDI Alignment of Piano Duets

KatieAnna E. Wolf
University of Minnesota
College of Science and Engineering
Minneapolis, Minnesota, USA
wolfx265@umn.edu

Elaine Chew
University of Southern California
Viterbi School of Engineering
Thornton School of Music
Los Angeles, California, USA
echew@usc.edu

Abstract: We present preliminary results from evaluations of MIDI-MIDI alignment of piano duet parts using Dynamic Programming. The end goal of this study is to create robust performance-to-score alignment for quantifying ensemble coordination in distributed performance.

In the study of performance, an advantage of using MIDI is that MIDI information contains actual note pitches, and precise onset times and durations. The disadvantage of using MIDI instruments is that it lacks the full range of expressivity present in acoustic instruments. Ensemble music presents unique challenges for sequence alignment, such as the existence of periods of silence when players wait their turn to play, during which no information is available on the passing of time when considering one part in isolation. Trills, which can have varying lengths and speeds from one performance to the next, and extremely fast notes, such as grace notes and cadenza-like runs, in piano music, present further problems for music alignment. In our particular example, long sequences of repeated notes or chords present similar challenges as that of periods of silence. Auditory delay that impacts ensemble coordination, particularly in distributed performance, increases errors (missed or wrong notes), and adds to the difficulty of the alignment.

We use an alignment technique based on Dynamic Programming that matches onsets in an expressionless score-generated MIDI file to a performed rendition of the piece captured in MIDI format. Onset times and pitches are extracted from the MIDI files using Toiviainen & Eerola's MIDI Toolbox (2004) implemented in Matlab. The feature used in each time window is the pitch set. A distance function is used to create a cost matrix which serves as input for Highfill's adaptation (2010) of Ellis' "simple" Dynamic Programming algorithm implementation (2003). We test two distance functions: Euclidean Distance and, for comparison, the measure used by Meron & Hirose (2001), which is the ratio of the cardinality of the intersection of the note sets and that of the union of the note sets.

Like Meron & Hirose, for each note in the score, we use its onset time to find the matched time slice in the expressive performance file, and search within a window for that note. If the note is found, then the alignment is deemed correct, and when it is not, we note that an error has occurred. We define the accuracy of an alignment as the number of notes correctly aligned over the number of notes in the piece. Inspired by Grosche, Müller, & Sapp's evaluation of MIDI-audio alignment for beat tracking (2010), we present details of our evaluation by systematically excluding different problematic note categories. Our interest extends beyond alignment of beats to the alignment of individual notes in the score. We adapt Grosche, Müller, & Sapp's categories to capture the special challenges posed by piano duet alignment: ostinatos (in particular, repeated notes or clusters of notes), boundary notes (notes just before or after two or more measures of rest), ornamented notes (trills, grace notes, and runs), target notes around ornamentations (notes after trills, after grace notes, and at the beginning of, or after, runs), and weak notes (passing tones and moving notes over a sustained note or notes).

To quantify piano ensemble coordination, it is important to determine an empirical onset time for notes that are synchronous in the score, but are almost never performed as such in practice. We test different local quantization techniques to determine empirical onsets for synchronous notes in the score. After alignment, each aligned onset is adjusted to the position of the (a) first, (b) loudest, (c) median, and (d) average onset inside a small time window surrounding that onset. We will present the results of the evaluations under these and other parameters using Sonic Visualiser 1.7.2 (2010).

Submitted July 9, 2010 for the ISMIR 2010 Late Breaking / Demo Session.

This work was supported in part by the Distributed Research Experiences for Undergraduates Program administered by the Committee on the Status of Women in Computing Research (CRA-W), and by the National Science Foundation (NSF) under Grant No. 0347988. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the CRA-W or NSF.