Human Computer Music Performance*

Roger B. Dannenberg

Carnegie Mellon University
Pittsburgh, PA 15213, USA
rbd@cs.cmu.edu

Abstract

Human Computer Music Performance (HCMP) is the study of music performance by live human performers and real-time computer-based performers. One goal of HCMP is to create a highly autonomous artificial performer that can fill the role of a human, especially in a popular music setting. This will require advances in automated music listening and understanding, new representations for music, techniques for music synchronization, real-time human-computer communication, music generation, sound synthesis, and sound diffusion. Thus, HCMP is an ideal framework to motivate and integrate advanced music research. In addition, HCMP has the potential to benefit millions of practicing musicians, both amateurs and professionals alike. The vision of HCMP, the problems that must be solved, and some recent progress are presented.

1998 ACM Subject Classification H.5.5 Sound and Music Computing, J.5 Arts and Humanities–Performing arts, H.5.1 Multimedia Information Systems

Keywords and phrases Interactive performance, music processing, music signals, music analysis, music synthesis, audio, score

Digital Object Identifier 10.4230/DFU.Vol3.11041.121

1 Introduction

Human Computer Music Performance (HCMP) is a new term intended to describe an emerging practice of creating computer music systems that can perform live music in association with human performers [11]. Interactive music performance itself is not new; however, even after decades of work in this area, examples of intelligent, competent, autonomous music performance by computer are rare. Moreover, some very general problems of music listening and generation need to be solved to enable a rich practice of HCMP. This contribution describes many problems of live music performance. It is hoped that researchers will be inspired to consider these problems and perhaps solve them.

Interactive music systems to date fall into several different categories. Probably, the most extensive work has been with experimental music where there are few traditions or constraints. This has freed creators from concerns of synchronization, harmonic structure, adherence to predetermined forms, etc. Instead, the focus can be on interactivity, gestural control, algorithmic composition, and new synthesis techniques, which have all advanced greatly over several decades [43, 46].

Another area of focus is score following and computer accompaniment [15]. These systems assume that music details are predetermined by the composer, and the main interactive task is synchronization. Typically, computer accompaniment systems have no "understanding" or representation of music theory, structure, or form, and there is no need to generate music other than to play predetermined notes or sounds. One might say computer accompaniment

* This work was partially supported by the National Science Foundation.

Multimodal Music Processing. *Dagstuhl Follow-Ups, Vol. 3.* ISBN 978-3-939897-37-8.
Editors: Meinard Müller, Masataka Goto, and Markus Schedl; pp. 121–134
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

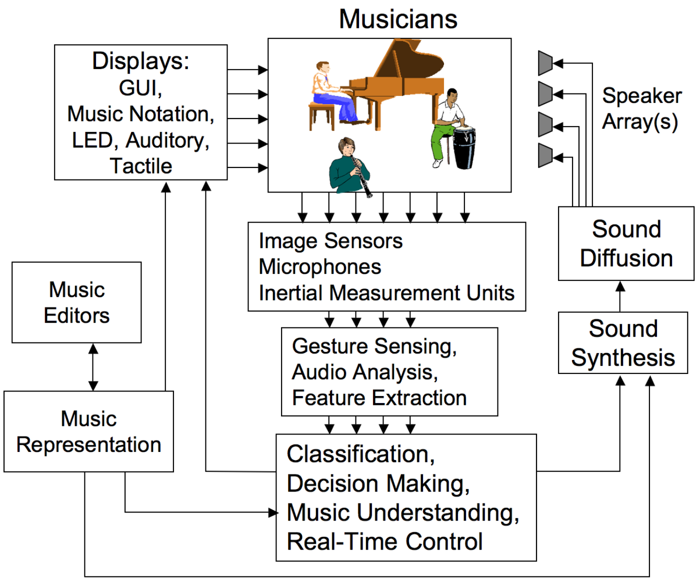

Figure 1 A high-level view of a Human Computer Music Performance (HCMP) system.

has been successful because all of the effort can be focused on two clear problems: (1) real-time alignment of a performance to a score, and (2) musically adjusting the playback of an accompaniment to synchronize to another player.

A limitation of computer accompaniment is that real-time alignment is not always useful or applicable to music synchronization. A common situation in popular music is that there is no detailed score to which a performance can be compared and aligned. For example, lead sheets may notate only chords. Even where a melody is notated in detail, it is often understood that the rhythms can be interpreted freely. Alignment to a score that is freely interpreted does not give very useful information about musical tempo and location. In these cases, a very different approach must be taken to synchronize musicians.

"Popular music" (for lack of a better term) is defined here to be music with a generally steady tempo, clear structures of sections, phrases, harmony, melody, and meter, and usually a substantial amount of improvisation. Synchronization in popular music tends to be based on beats and measures rather than on a score. The score is likely to be imprecise (as in a lead sheet that only specifies chords and melody), and even when the score suggests every note to play, musicians are usually free to interpret the music, adding their own chord voicings, strumming patterns, syncopations, etc. Because of the difficulty of synchronizing machines to humans, it is common for humans to synchronize to machines or to the playback of fixed media as in Karaoke performance, street musicians playing along with backing tracks, and practicing with Music Minus [36] recordings.

At present, HCMP is more of a vision than a practice. The vision is for autonomous computer systems to play the role of a skilled musician in a live performance of popular music.[1] Figure 1 gives a high-level view of a complete HCMP system. Musicians are sensed not only by machine listening but also through a variety of sensors and interfaces. Musical

[1] The term HCMP is really applicable to all forms of live music performance involving humans and computers. If the term is used more broadly, then we may need more specific terms, e.g. HCMP_{PM} for popular music, HCMP_{SF} for synchronization by score following, etc. I will simply use HCMP for now.

decision making relates real-time music performance to music representations that could be very specific or only musical sketches. Music is generated, synthesized, and diffused via loudspeakers into the performance space. Non-audio feedback in the form of displays, tactile feedback, and music notation are also generated to communicate with the live performers. Editors are also available to create and alter music for the HCMP system. Realizing the vision of HCMP will require solutions to a number of problems:

- Synchronization in beat-based music is largely unexplored from the standpoint of implementing systems with musical competence. Of course, automatic beat tracking has been studied, but highly reliable and robust systems do not exist. Studies of human synchronization in music are found in the literature of performance analysis, for example [41].
- Communication among musicians is especially important in popular music where the score may only be a sketch and where performers are generally free to alter the form of the music in mid-performance.
- Musicians plan performances based on very abstract representations of the music. For example, "I'll solo and you come in on the bridge" is almost a complete recipe for performing a ballad, but this is possible only because the musicians have shared conventions for describing music structure and organizing performances.
- Musicians transform some representation of music into live sound. This may involve the composition of parts, e.g. writing a bass line given a chord progression. Once notes and phrases are determined, they must be synthesized musically or perhaps performed acoustically by a robot. Finally, sounds must be diffused into the performance space, ideally in a way that conveys the impression of live performance rather than the mere playback of a recording.

After a short discussion of related work, these problems are considered one-by-one in the following sections. We conclude with a discussion of the possible impact HCMP research can have on the future of music.

2 Related Work

As mentioned above, most of the work to date on interactive computer music addresses problems of experimental contemporary art music and Western art music, but ignores the problems of popular music or music with a steady tempo, where scores are not so helpful and where timing must be very precise. In the commercial world, Ableton Live [1] provides a powerful interface for beat-based music production and control, and it provides some real-time time-stretching and tempo adjustment capabilities, but it is not meant to function as a virtual musician. Robertson and Plumbley's B-Keeper system [42] extends Ableton Live with a real-time beat tracker and user interface so that a user can synchronize music to a live drummer. This system implements some components of an HCMP system, but does not address other issues.

Conducting systems [5, 8, 13, 29] are closely related to HCMP, and differ mainly in that they assume a dedicated person (the conductor) to give commands to the computer. Conducting is certainly an interesting way to synchronize computer performers with humans in a live performance. Conducting systems have been used in public performances, including some controversial performances of traditional works where electronics was used in place of an acoustic orchestra [19].

In some conducting systems, all of the music is assumed to be generated by the computer, so cues and synchronization are not critical issues. Many conducting systems have focused on issues of adjusting tempo in real time and performing real-time time stretching of audio and video in live interactive performance of classical music.

On the other hand, popular music performance with human musicians raises the issues of accurate synchronization, and how to handle cuts, repeats, and other changes in form. Another issue is that conducting systems require a human conductor. Even if there is already a conductor, it is common to add another conductor specifically to operate the computer system. This is important because any technical "solution" that requires the full-time efforts of a person has to be compared to the possibility of simply adding another human musician to the existing ensemble. In contrast to a conducting system, an autonomous virtual musician seems to be a better solution for a small group playing popular music.

3 Beat Tracking, Tempo, and Synchronization

The irony of working with "steady beat" music is that in live performance, the tempo is never truly steady. Variations of 5 to 10 percent over time periods on the order of one minute are to be expected, and fairly sudden tempo changes are common as well. In spite of this variation, "steady beat" music is, for the most part, very steady, and the predictability of beat times is crucial to synchronization, given that all the parts may be improvised and therefore unpredictable. HCMP systems need to identify beats accurately and reliably. This might be achieved with a combination of automatic beat tracking software [7, 16, 17, 23, 24, 32, 35] and gestural sensors such as foot pedals or accelerometers.

Automation is desirable but tends to be not so reliable in two ways. First, automatic beat trackers often make serious mistakes, losing track of the beat altogether. Although the literature often implies that beat tracking is largely solved, even state-of-the-art beat trackers fail too often to be used in live performance. The second problem is that automatic beat tracking precision is not high. One might expect that when beat trackers work, they are synchronized to audio features such as snare drum hits, which should be easy to detect, reliable, and precisely timed. In reality, music audio often contains events that are slightly offset from the true beat times and which result in inaccuracies. It seems that humans and automatic beat trackers are using very different processes to identify beats, and these processes are not always consistent.

Human input through tapping is more robust that beat-tracking, but tapping can be a distraction to musicians. Also, musicians who are distracted by performance tasks can tap with inadvertently large skews between the tap times and the true beat times. As with automatic beat tracking, there are precision problems with tapping, especially foot tapping, which otherwise is one of the most reliable and least obtrusive ways of getting beat information from live performers.

In addition to identifying beats, HCMP systems need to know how beats relate to the overall music structure. An important level of structure is the measure or bar. These groups of 4 beats (typically) are the points of transitions such as chord changes and phrase beginnings. By determining measure boundaries, HCMP systems can better interpret cues or signals from humans. These cues are often ambiguous at the beat level but usually refer to the nearest measure boundary. Robustly detecting measure boundaries is an interesting real-time music analysis problem [27, 38].

4 Human Computer Communication

While computer accompaniment systems and beat trackers focus on extracting information from music audio, much of the interaction in a music performance is external to the audio channel. Musicians give visual signals, make eye contact, and use body gestures to communicate music information, including cues for synchronization. Not all signals are explicit. For example, as a trumpet player, I can usually tell when a fellow player is about to play by noticing when the trumpet is lifted and when the player takes a breath. I may not even be consciously aware of this communication, but it exerts a strong influence that may help me to take corrective action if I am not synchronized to the other players.

Similarly, HCMP systems should develop simple and natural communication channels between computer and human musicians. In fact HCMP is intended to sound a bit like HCI–Human Computer Interaction–and HCMP research needs to build upon HCI techniques and approaches [40]. This is a rich area for research, especially given the many possibilities of sensors and computer processing. We need to explore real-time interaction techniques for performing musicians as well as real-time displays that allow computers to communicate with and give cues to humans.

Nicolas Gold and I [22] have identified multiple classes of cues in an effort to describe musical communications more systematically. A *Static Score Position Cue* communicates the current position in the score to correct synchronization problems. An *Intention Cue* is used to indicate the direction of the performance when there are options. For example, "this is the last time we will repeat this section." A *Voicing/Arrangement Cue* is used to modify the performance, for example by telling a player to begin playing, to play louder, or to play more notes.

We have developed a music notation-based interface for HCMP in my lab [30]. (See Figure 2.) The music notation system, inspired by earlier work [10], is bi-directional: The computer can display its location by highlighting bar lines in the music to confirm to the human that it is in the right location. On the other hand, the human can touch or click on locations in the score to give cues or to tell the computer where to begin in a rehearsal.

Other modes of interaction are also possible. For example, we have developed a small touch sensor that can be worn on the finger and used to give cues while playing a musical instrument. Nicolas Gold wrote software to interpret several different free hand cues using a Kinect [33]. I have used foot tapping at half tempo (cut time) to indicate tempo and 4 taps at full tempo to cue the beginning of some music. The possibilities seem endless.

5 Music Structure

Music structure and representation has received much attention, but HCMP seems to raise some unique questions. The main idea is that music often includes repetition and hierarchical structure. For example, a popular song form can be described as "AABA," indicating that the first section (usually 8 measures) is repeated, followed by a contrasting "B" part, and then the "A" part is repeated again at the end. In practice, sections are usually varied slightly from one instance to the next. An interesting challenge for music processing is to detect music structure automatically from audio [2, 14, 39] or symbolic scores [28].

Popular music scores exist in multiple forms, and coordinating these somewhat informal objects systematically will require careful design of models, representations, and interfaces [22]. The typical musical score uses repeat signs and other constructs to make the notation more compact. Often, there are exceptions where the music is played differently on each repetition. Sometimes, the exceptions are handled in a standard way, such as first and second ending notation, but often there are informal annotations, e.g. "Play 1x only." We call this the *static* score, in analogy to static computer programs.

When a static score is "executed" or unfolded to create a linear sequence of events, the fact that repetitions are different and occur at different times can be represented directly,

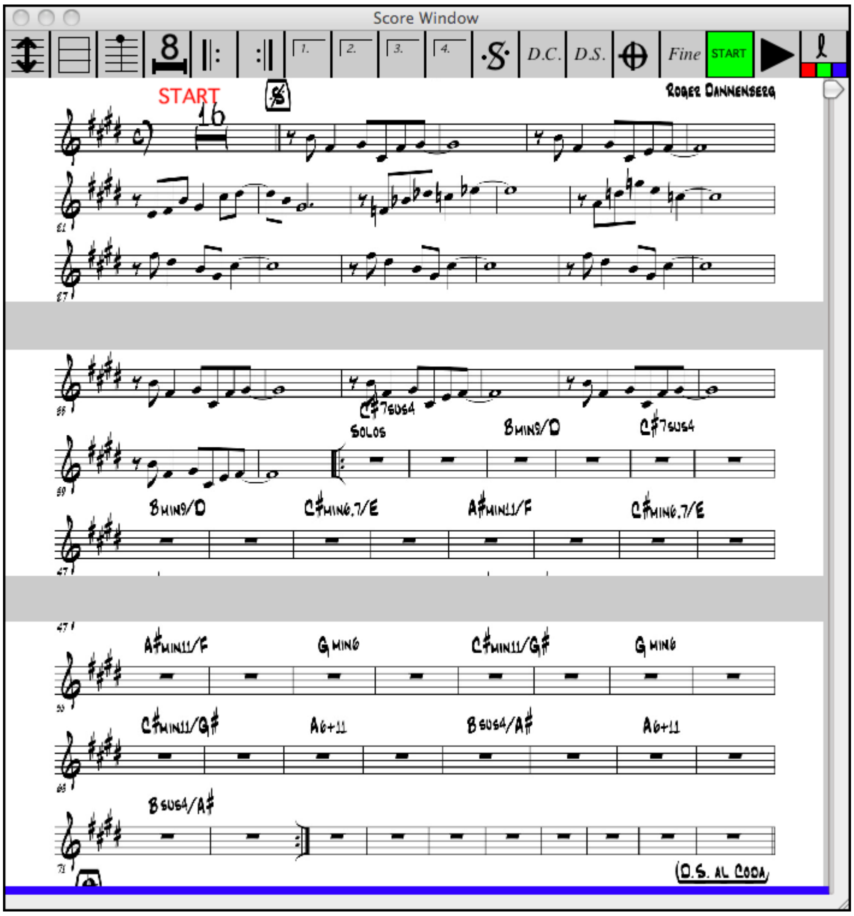

Figure 2 A bi-directional interactive music interface for HCMP. The computer can scroll music and highlight its current position. The human can point to locations to give cues, correct synchronization problems, or indicate starting locations in rehearsals.

albeit with greater redundancy. We call this the *dynamic score* in analogy to the dynamic (run-time) execution sequence of a program. The mapping between static scores and dynamic scores is complicated by non-determinism. For example, a repeat may be marked "ad lib," meaning that the number of repetitions is to be determined at performance time. Thus, the dynamic score cannot be fully determined until performance time.

However, dynamic scores are not just ephemeral traces of a performance. An audio recording corresponds to the dynamic score as does a MIDI sequence. If a performance consists of humans reading static scores and an HCMP system playing from a MIDI file and an audio file, then clearly the static and dynamic representations must be reconciled. See [21] for related work.

In popular music, scores are often informal, and musicians often create informal plans that do not match the implied plan of the score. For example, musicians might decide to play intro, verse, chorus, chorus, ending, even if the score shows a second verse. In popular music, it is accepted practice to alter the structure in this way. Musicians might even decide to change the key of the second chorus. We call these informal plans "arrangements." An HCMP system must be able to access information from a static score, a dynamic score, and an arrangement in order to make a performance plan that is consistent with the intentions of human musicians. It must be easy for human musicians to make and communicate arrangements in a few seconds or even during the performance, because these decisions are often made on the spot during a performance. Building simple tools to manage music where these fairly abstract concepts come into play is also a challenge.

6 Music Generation, Synthesis, and Diffusion

There are many ways that computer musicians can actually produce sound. The simplest technique is to use MIDI scores and conventional synthesizers. This approach may work very well in certain situations, but it assumes there is a detailed score including dynamics and expressive timing, and that the instrument can be synthesized adequately. An alternative is to play back prerecorded audio using time-stretching techniques to adjust the tempo of the recording and synchronize it to the live performers [12]. This works well, but it requires a lot of time and effort to prepare and it assumes that live musicians are available to perform the necessary parts in the recording session. Current synthesis methods for many instruments are not very convincing, and even good synthesis methods often require such careful control that satisfactory results are very hard to obtain. Progress is being made in research systems, e.g. [18, 37], and commercial systems, e.g. [31, 34, 44], but HCMP could benefit from new research in synthesis methods. Ideally, one would like to render a score into a convincing performance including musical phrasing, stylistically appropriate timing and articulation, dynamics, and vibrato.

A common task for a popular music performer is to play according to a "lead sheet" or "chord chart," which specifies the key, harmony and structure of the music (and sometimes the melody), but few if any other details. From this information, a drummer can create an appropriate rhythm that matches the structure of the song, a bass player can provide a rhythmic and harmonic foundation, a keyboard or guitar player can play chords according to the harmony, and other musicians can harmonize the melody or improvise a solo. Writing a new song might be considered a highly creative and difficult task, but creating a bass part or playing piano to accompany a singer is a routine task for a working musician. In fact, most musicians can create very musical parts without errors in real time as they read a lead sheet. In the field of popular music performance, many musicians are actually more comfortable improvising parts from a lead sheet than reading conventional music where every note is indicated explicitly. Thus, creating musical parts from a sketch such as a lead sheet is a fundamental skill that should be implemented by an HCMP system. This is especially important when human performers do not already play the instrument to be played by HCMP and therefore lack the skill or experience to compose the part.

Programs such as Band-in-a-Box [26] perform the music-from-lead-sheet task already, but do not give the user too much control over music generation. Instead, Band-in-a-Box offers a wide variety of styles from which the user can select. It seems that research into machine learning, musical analogy, music similarity, and models of musical style can lead to more flexible and controllable music generation.

Finally, one of the problems with computer-generated sound is diffusion into an acoustic space. The one-dimensional audio signal must be converted to three-dimensional sound waves through loudspeakers, which impart their own audible radiation characteristics onto the sound. As computer and digital audio equipment prices have fallen, there has been much interest in using many audio channels to drive arrays of speakers to improve and control the diffusion of sound in two or three dimensions. Examples include linear speaker arrays to produce controlled wavefronts [4, 6], spherical speaker arrays to simulate sound sources with frequency-dependent radiation patterns [3, 9], and our own work on convolution-based stereo panning and placement [25]. The challenge for HCMP is to completely integrate computer performance with live acoustic instruments. The techniques will depend upon the situation, but there is a need for improvements in sound reinforcement with implications for all audio interfaces.

7 Example

One substantial performance [12] using HCMP has been presented at Carnegie Mellon University. I worked with the student jazz band, with assistance from David Pellow, the director, and John Wilson, an arranger. We set out to create a string orchestra that could play along with a live jazz band. Our musical goal was that the strings plus jazz band sound as good as possible. We decided to emphasize practical considerations and reliability over cutting-edge research, and our experience has helped to formulate some of the problems and approaches suggested above.

John Wilson was commissioned to write for jazz band and strings. We decided to structure the piece so that the string players do not play continuously, but instead play during a number of segments of the music. The human rhythm section (drums, bass, and piano) plays continuously throughout the piece. This allowed for interplay between the band and the strings. It also had functional purposes: Because the strings were silent at many times, each entrance could be cued separately. If anything went wrong, there would soon be an opportunity to make another entrance. In addition, the sectional nature allowed for efficient recording and editing.

To the computer, the string parts are just sounds that need to be cued to begin on a particular beat and synchronized to the following beats. When each sound ends, the system prepares to play the next. Synchronization is handled in the simplest way imaginable. A foot pedal is used to tap beats (in cut time, about 85 taps per minute) to establish the tempo and throughout the performance. A small keyboard is used to cue entrances.

One thing we learned is that some new listening skills are required. In one rehearsal, the tapper (a skilled percussionist) started listening to the string section under control and naturally started to tap along with the strings rather than the band. Once the band was ignored, the tapper and strings started to drift away from the live band. As soon as the strings were obviously out of synchronization, the tapper resynchronized to the live band, but I had written code to reject “spurious” taps that fell more than 30% of a beat from the expected beat time, and the corrective taps were ignored. (Since then, we have disabled the “spurious” tap rejection feature.)

Sound generation is based on the pitch-synchronous, overlap-add (PSOLA) [45] approach to time stretching. The requirement for real-time performance, continuous (every tap) update, latency compensation, and synchronization across multiple (20) channels led to some innovative implementation details. Most PSOLA systems are designed to time stretch by a given factor over a given time span. PSOLA works by inserting or deleting whole pitch periods, thus the operation of PSOLA is not really continuous stretching but an approximation that is quantized to pitch periods. We have 20 separate channels, each with a single instrument and its own set of pitch periods. Due to the quantization, it is difficult to predict the exact duration of input audio that will be consumed, and there is the possibility that stretched tracks will lose synchronization with the accumulation of quantization errors. To ensure that all tracks remain synchronized, we use a feedback mechanism: The overall control system adjusts the global stretch factor so that the mean audio file position will

Figure 3 An HCMP performance. The live jazz ensemble is complemented by a 20-piece virtual string orchestra played over an array of loudspeakers visible behind the band.

synchronize with the live band. Then, the per-track stretch factor is adjusted slightly for each track to drive the track's audio file position toward the global mean.

Finally, sound diffusion is based on multiple (8) speaker systems arranged across the stage (see Figure 3). Each of the 20 input channels represents one close-miked string (violin, viola, or cello). Each instrument channel is directed to only one speaker. Rather than a homogenized orchestra sound spread across many speakers, we have individual instrument sounds radiating from multiple locations and mixing in the room as with an acoustic ensemble. The results were so convincing that an audio engineer wanted to know what reverberation plug-in we used for the recordings, yet the recordings were actually dry and all sense of "stereo" and reverberation resulted from the diffusion scheme.

8 Conclusions

From a scientific standpoint, Human Computer Music Performance (HCMP) offers a framework to organize, motivate, and coordinate an array of interesting research efforts. It suggests that we study music and music performance from many points of view, developing new techniques for sensing music beats, tempo, and structure, as well as new ways for musicians to communicate music intentions, especially to computer-based performers. The fundamental problems underlying HCMP are general problems of Music Understanding, thus there are broad implications. As the "Multimodal Music Processing" theme of this book suggests, this work is truly multimodal, dealing with various levels of discrete and symbolic scores, music performance data such as MIDI, music audio, graphical displays, gesture sensors, and other forms of musical communication. The models and analysis techniques introduced will have applications in other music-related studies such as music information retrieval, music theory, and music cognition.

Digital sound synthesis has been an object of study for half a century, and has connections to auditory perception, acoustics, digital signal processing, speech synthesis, and mathematics. HCMP challenges us to investigate new techniques for time-stretching and pitch shifting as well as to consider the importance of sound diffusion in the perception of music synthesis quality.

More broadly, HCMP addresses complex, real-time cooperative tasks. New interfaces are needed to coordinate computers and humans with a minimal amount of explicit or manual control. This could have implications for other human-computer interaction scenarios such as driving and piloting, directing disaster relief, or complex mission control where tasks must be delegated and coordinated. HCMP problems suggest an integrated approach that combines machine learning with human factors studies to create reliable interfaces; advances in this area should have many applications beyond music.

Music making is practiced in the majority of households in the United States. The National Association of Music Merchants reported that industry retail sales in 2006 were about 8 billion dollars in the U.S. This includes sales of over 5 million musical instruments, but does not include music education, music recordings, or music performance. Thus, the potential societal impact of effective new music technology is enormous. HCMP seems to be an application area where recent advances in music understanding and music information processing can be leveraged to benefit millions of people. Producing results that are extremely practical and useful is not just an altruistic project for researchers. By integrating academic research to create a practice of popular music-making, the research community stands to gain greater recognition and support from society.

Another motivation for research in Human Computer Music Performance is purely artistic. One could criticize HCMP as an effort to further reduce and automate popular music, which is already formulaic. Would it not be better to devote efforts to experimental music and new art forms? My hope is to leverage the conventions, opportunities, and sheer numbers in popular music to obtain a widespread practice of HCMP. I am convinced that if this succeeds, at least a few artists in a million will invent some truly creative uses for HCMP technology that transcend existing musical practice. Thus, HCMP could be an important path by which technology shapes the future of music.

9 Acknowledgements

Thanks to Ryan Calorus, who implemented our first experimental music display, Nicolas Gold for valuable discussions, and my students Dalong Cheng, Zeyu Jin, Dawen Liang, Jiuqiang Tang, and Gus Xia. Our first performance system and the music display work were supported by Microsoft Research and the Carnegie Mellon School of Music. Zplane kindly contributed their high-quality audio time-stretching library [20] for our use, and David Pellow, Riccardo Shulz, and John Wilson made essential musical contributions to our performance with the Carnegie Mellon Jazz Ensemble. Current work is supported by the National Science Foundation under Grant No. 0855958.

References

1 Ableton AG. *Ableton Reference Manual*, 2010.

2 Jean-Julien Aucouturier and Mark Sandler. Finding repeating patterns in acoustic musical signals: Applications for audio thumbnailing. In *AES22 International Conference on Virtual, Synthetic and Entertainment Audio*, pages 412–421. Audio Engineering Society, 2002.

3 Rimas Avizienis, Adrian Freed, Peter Kassakian, and David Wessel. A compact 120 independent element spherical loudspeaker array with programmable radiation patterns. In *Proceedings of the AES 120th Convention*, 2006. Paper No. 6783.

4 Marije A. J. Baalman. Application of wave field synthesis in electronic music and sound installations. In *The ICMC 2004 Proceedings*, pages 692–698, San Francisco, 2004. The International Computer Music Association.

5 Takashi Baba, Mitsuyo Hashida, and Haruhiro Katayose. “VirtualPhilharmony”: A conducting system with heuristics of conducting an orchestra. In *Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010)*, pages 263–270. ACM Press, 2010.

6 A. J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America*, 93(5):2764–2778, 1993.

7 Paul Brossier. *Automatic Annotation of Musical Audio for Interactive Applications*. PhD thesis, Department of Electronic Engineering, Queen Mary, University of London, 2006.

8 Bernd Bruegge, Christoph Teschner, Peter Lachenmaier, Eva Fenzi, Dominik Schmidt, and Simon Bierbaum. Pinocchio: conducting a virtual symphony orchestra. In *ACE '07 Proceedings of the international conference on Advances in computer entertainment technology*, pages 294–295. ACM, 2007.

9 Perry R. Cook, Georg Essl, Georgos Tzanetakis, and Dan Trueman. N >> 2: Multi-speaker display systems for virtual reality and spatial audio projection. In *Proceedings of the International Conference on Auditory Display (ICAD)*, 1998.

10 David Damm, Christian Fremerey, Verena Thomas, Michael Clausen, Frank Kurth, and Meinard Müller. A digital library framework for heterogeneous music collections—from document acquisition to cross-modal interaction. *International Journal on Digital Libraries: Special Issue on Music Digital Libraries*, 2011, to appear.

11 Roger B. Dannenberg. New interfaces for popular music performance. In *Seventh International Conference on New Interfaces for Musical Expression: NIME 2007*, pages 130–135, New York, NY, June 2007. New York Univ., ACM Press.

12 Roger B. Dannenberg. A virtual orchestra for human-computer music performance. In *Proceedings of the 2011 International Computer Music Conference*, pages 185–188. The International Computer Music Association, 2011.

13 Roger B. Dannenberg and Ken Bookstein. Practical aspects of a midi conducting program. In *Proceedings of the 1991 International Computer Music Conference*, pages 537–540. International Computer Music Association, October 1991.

14 Roger B. Dannenberg and Masataka Goto. *Music Structure Analysis from Acoustic Signals*, volume 1, pages 305–331. Springer Verlag, 2009.

15 Roger B. Dannenberg and Christopher Raphael. Music score alignment and computer accompaniment. *Communications of the ACM*, 49(8):38–43, August 2006.

16 Matthew E. P. Davies and Mark D. Plumbley. A spectral difference approach to downbeat extraction in musical audio. In *Proceedings of the 14th European Signal Processing Conference (EUSIPCO 2006)*, 2006.

17 Daniel P. W. Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, 2007.

18 Gianpaolo Evangelista and Fredrik Eckerholm. Player-instrument interaction models for digital waveguide synthesis of guitar: Touch and collisions. *IEEE Transactions on Audio, Speech and Language Processing*, 18(4):822–832, 2010.

19 Shirley Fleming. The virtual orchestra. *American Record Guide*, 66(6), Nov./Dec. 2003.

20 Tim Flohrer. *Elastique 2.0 SDK Documentation*. zplane.development, 2007.

21 Christian Fremerey, Meinard Müller, and Michael Clausen. Handling repeats and jumps in score-performance synchronization. In *Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR)*, pages 243–248, Utrecht, Netherlands, 2010.

22 Nicolas Gold and Roger B. Dannenberg. A reference architecture and score representation for popular music human-computer music performance systems. In *Proceedings of the 2011 International Conference on New Interfaces for Musical Expression (NIME11)*, 2011.

23 Masataka Goto. An audio-based real-time beat tracking system for music with or without drum sounds. *Journal of New Music Research*, 30(2):159–171, 2001.

24 Peter Grosche and Meinard Müller. Extracting predominant local pulse information from music recordings. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(6):1688–1701, 2011.

25 William D. Haines, Jesse R. Vernon, Roger B. Dannenberg, and Peter Driessen. Placement of sound sources in the stereo field using measured room impulse responses. In *Proceedings of the 2007 International Computer Music Conference, Volume I*, pages I–496–499, San Francisco, August 2007. The International Computer Music Association.

26 PG Music Inc. Band-in-a-box 2012. `http://www.pgmusic.com` (retrieved 07.03.2012), 2012.

27 Anssi P. Klapuri, Antti J. Eronen, and Jaakko T. Astola. Analysis of the meter of acoustic musical signals. *IEEE Transactions on Audio, Speech and Language Processing*, 14(1):342–355, 2006.

28 Olivier Lartillot. A musical pattern discovery system founded on a modeling of listening strategies. *Computer Music Journal*, 28(3):53–67, 2004.

29 Eric Lee, Thorsten Karrer, and Jan Borchers. Toward a framework for interactive systems to conduct digital audio and video streams. *Computer Music Journal*, 1(30):21–36, Spring 2006.

30 Dawen Liang, Guangyu Xia, and Roger B. Dannenberg. A framework for coordination and synchronization of media. In *Proceedings of the 2011 International Conference on New Interfaces for Musical Expression (NIME11)*, pages 167–172, 2011.

31 Eric Lindemann. Synful. `http://www.synful.com` (retrieved 07.03.2012), 2012.

32 M. F. McKinney, D. Moelants, M. E. P. Davies, and A. Klapuri. Evaluation of audio beat tracking and music tempo extraction algorithms. *Journal of New Music Research*, 36(1):1–16, 2007.

33 Microsoft. Xbox 360 + kinect. `http://www.xbox.com/kinect` (retrieved 06.03.2012), 2012.

34 Modartt. Pianoteq - product catalog. `http://www.pianoteq.com/catalog` (retrieved 07.03.2012), 2012.

35 João Lobato Oliveira, Fabien Gouyon, Luis Gustavo Martins, and Luis Paulo Reis. Ibt: A real-time tempo and beat tracking system. In *Proceedings of the 11th International Conference on Music Information Retrieval*, pages 291–296, 2010.

36 Music Minus One. Music minus one home page. `http://www.musicminusone.com` (retrieved 06.03.2012), 2012.

37 Jyri Pakarinen. Physical modeling of flageolet tones in string instruments. In *Proceedings of the 13th European Signal Processing Conference*, pages 4–8, 2005.

38 Hélène Papadopoulos and Geoffroy Peeters. Joint estimation of chords and downbeats from an audio signal. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(1):138–152, 2011.

39 Jouni Paulus, Meinard Müller, and Anssi Klapuri. Audio-based music structure analysis. In *Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR)*, pages 625–636, Utrecht, Netherlands, 2010.

40 Jenny Preece, Yvonne Rogers, Helen Sharp, David Benyon, Simon Holland, and Tom Carey. *Human-Computer Interaction: Concepts and Design.* Addison-Wesley Pub. Co., 1994.

41 Bruno H. Repp, Justin London, and Peter E. Keller. Production and synchronization of uneven rhythms at fast tempi. *Music Perception*, 23:61–78, 2005.

42 Andrew Robertson and Mark Plumbley. B-keeper: A beat-tracker for live performance. In *New Interfaces for Musical Expression*, pages 234–237, 2007.

43 Robert Rowe. *Interactive Music Systems.* MIT Press, Cambridge, MA, 1993.

44 Samplemodeling. Sample modeling products. `http://www.samplemodeling.com/en/products.php` (retrieved 07.03.2012), 2012.

45 Norbert Schnell, Geoffroy Peeters, Serge Lemouton, Philippe Manoury, and Xavier Rodet. Synthesizing a choir in real-time using pitch synchronous overlap add (PSOLA). In *Proceedings of the 2000 International Computer Music Conference*, pages 102–108, 2000.

46 Tod Winkler. *Composing Interactive Music: Techniques and Ideas Using Max*. MIT Press, Cambridge, MA, 2001.