A Vision of Creative Computation in Music Performance

Roger B. Dannenberg
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213 USA
rbd@cs.cmu.edu

Abstract

Human Computer Music Performance (HCMP) is the integration of computer performers into live popular music. At present, HCMP exists only in very limited forms, due to a lack of understanding of how computer performers might operate in the context of live music and a lack of supporting research and technology. The present work examines recent experimental systems as well as existing music performance practice to envision a future performance practice that involves computers as musicians. The resulting findings provide motivation as well as specific research objectives that will enable new creative practice in music.

Introduction

Sound and music computing research has focused attention mainly in two directions: "high art" and commercial music from the recording industry. Largely ignored is the live performance of popular music including rock, jazz, and folk music. While the practice of *popular* music is not currently a hot topic for music technology research, it is arguably the dominant form of live music. In a recent weekly listing of concerts in Pittsburgh, there are 24 "classical" concerts, 1 experimental/electro-acoustic performance, and 98 listings for rock, jazz, open stage, and acoustic music. By examining the features of popular music practice, we find many commonalities across a diverse array of musics, including rock, jazz, folk, music theater, church music, choral music, and others. Live music in all of these popular forms offers a wealth of opportunities for computing and music processing research. I call the integration of computers as performers into popular live music performance practice "Human-Computer Music Performance" (HCMP).

Technologies such as digital audio recording and music synthesizers have changed the musical landscape dramatically over the last few decades. I believe an even more radical and creative musical revolution is in progress, where computers become *performers* rather than merely *instruments*. Because this revolution will involve new musical forms and new tasks (not just the automation of known ones), it is difficult to imagine and predict.

HCMP will be most interesting when computers exhibit human-level musical performance, but this is such a giant advance over current capabilities and understandings that it offers little guidance for HCMP research in the short term. An alternative is to envision a future of HCMP based on realistic assumptions of machine intelligence. Thus, an important initial step in HCMP research is to imagine how HCMP systems will operate. A clear vision of HCMP will motivate research to make it happen.

This is a position paper that poses a "grand challenge" for creative computing in music. Rather than stating a vague problem with no path to a solution, I hope to make the vision concrete enough to pose specific problems and research directions. I will consider some intermediate steps and approaches toward the ultimate goal. The results so far are not solutions themselves, but objectives that define problems and motivate their solutions.

There is an apparent contradiction in saying, on the one hand, that my goal is to develop "new musical forms and new tasks," but on the other hand, that I wish to address existing forms of live popular music. Since interactive computer music, aside from electronic instruments, is rarely used in popular music, there is a huge opportunity for a new and creative synthesis of ideas from which new forms and new tasks will emerge. This synthesis will only occur if musicians are first motivated to integrate more creative computing into their current practice. Thus, my strategy is to bridge the gaps that limit the use of computing in live popular music. New technologies will create opportunities for more radical transformations and artistic innovation.

The next section describes different modes of interactive computer music. Then current work in HCMP is presented. "The Vision of HCMP" describes requirements for future systems and makes specific predictions as to how these will be met. The paper ends with "Future Work" and "Conclusions.

Human-Computer Music Performance

Computers have been used in music performance for many years, so before going further, we should discuss HCMP and explain how this differs from current practice in computer music. (See Table 1.) The most common use of computing in music performance is through *computer* *instruments*, typically keyboards. These, and other electronic instruments, are essentially substitutes for traditional instruments and rely upon human musicians for their control and coordination with other musicians. Many composers of *interactive contemporary art music* use computers to generate music in real time, often in response to live performers. These works typically take advantage of contemporary trends toward atonality and absence of a metrical pulse, which simplifies the problems of machine listening and synchronization. Alternatively, the practice of *computer accompaniment* solves the synchronization problem by assuming a pre-determined score (music notation) to be played expressively by the performer while the computer follows the performer in the score and synchronizes an accompaniment. Other solutions to the synchronization of computers and humans include using *fixed media* with "click tracks" that performers can hear through headphones (with minimal interaction), and *conducting systems*, where a human essentially taps a click track for the computer.

Table 1. Interactive Music Major Threads

Computer Instruments	Direct physical interaction with virtual instruments: digital keyboards, drums, etc.
Interactive Contemporary Art Music	Composed interactions; often unconstrained by traditional harmony or rhythm. Digital audio effects and transformations of live performance.
Computer Accompaniment	Assumes traditional score; score following synchronizes computer to live performer.
Fixed Media	Many musical styles and formats. Live performers synchronize to fixed recording.
Conducting Systems	Synchronize live computer performance by tapping or gesturing beats. Best with "expressive" traditional/classical music.
HCMP	Assumes mostly steady tempo and synchronization to beats, measures, and sections. Compatible with improvisation at all levels.

In contrast, *HCMP* is aimed toward "common practice" music where performers improvise to varying degrees, where tempo is fairly steady, and where the structure of the music may be determined spontaneously during a performance. The "improvisation" assumption says that performers make musical choices ranging from strumming styles and bass lines to full-blown jazz improvisation. Since there is no pre-determined note-level description of the performance, musicians must synchronize on the basis of a more-or-less implied temporal structure of beats, measures, and sections. We also assume that performers can take liberties with the music, adding an introduction, extending a song with an instrumental solo, etc. Often, these changes are determined during the performance, so performers (human or computer) must be flexible enough to recognize and adapt to these changes. As mentioned in the introduction, HCMP addresses the most commonly performed musical styles, including rock, folk, and jazz.

HCMP Systems to Date

Before describing a vision of future systems, let us look at three systems that have already been built. These are two systems for performing with live jazz musicians and one system for electronic display of music notation.

Virtual Orchestras, Virtual Music Players

The first system is a virtual string orchestra that performs with a live jazz band (Dannenberg, 2011). This system uses studio recordings of acoustic instruments (violins, violas, and cellos) that are synchronized to the live band using audio time-stretching software based on PSOLA (Schnell *et al.* 2000). This produces high-quality output as long as the source sound is a single, periodic tone. Therefore, we recorded each string part separately, resulting in a 20-channel audio file, and we stretch each channel independently to form a variable-tempo, 20-piece orchestra.

To synchronize with the live band, a band member taps a foot pedal in time with the music, and some simple outlier rejection and linear regression software processes the data to form an estimate of the current and future beat position as a function of time. The virtual orchestra does not play continuously, but instead has about 10 separate entrances. Most entrances are cued by pressing a key in time with the music, allowing the system to resynchronize if anything goes wrong or if a soloist decides to play some extra measures.

The strings are mixed into 8 audio channels, which are played by 8 high-quality speakers arranged spatially to give the sound of a full ensemble playing in a concert hall. (See Figure 1.) This helps the strings match the 3-dimensional presence of the live musicians.

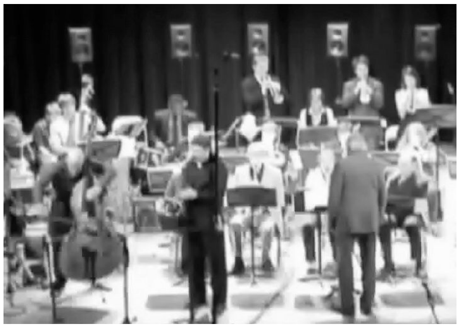

Figure 1. Jazz band performance with virtual strings (played by speakers in background).

The second system is a somewhat scaled-down version of the first. Rather than carefully recording and editing music for the virtual player, this version uses MIDI. In this particular case, the part would otherwise be played on a MIDI keyboard, so MIDI is not a limitation. This system uses a foot-tapping input for tempo acquisition, and the foot pedal is also used to cue entrances. Sections are played in sequence, but the operator can override the order by clicking a button on a computer screen. The computer part is mostly eighth-note arpeggios over a rock beat, making precise synchronization very important.

Electronic Music Display

One of the problems of HCMP is communication among both computer and human performers. Since musicians often read music during a performance, a computer-based visual display of music provides an interesting potential for a real-time, two-way musical communication channel.

To ease the transition from traditional paper, we assume printed music will be digitally photographed or scanned. Our software loads image files and offers some simple editing where users can mark logical page boundaries (normally between systems or staves).

Digital music displays are not new, but the concept of a music display as a musical communication device has received very little attention beyond the concept that a conductor could remotely ensure that all musicians are viewing the same music (Connick 2002). One of the areas we have been investigating is the mapping between the "static score" such as printed music and the "dynamic score" which is represented by recorded music or a live performance. The static score includes instructions to repeat sections, jump back and play the beginning again, etc., whereas the dynamic score is in some sense an instantiation or unfolding of the static score. These concepts are important because if a musician points to some notation, or if the computer highlights a score location, that location might represent the first, second, or third repetition in terms of the dynamic score. Mechanisms are needed to disambiguate static score locations.

Another focus of our work is page layout and page turning. We have shown, for example, that if a musician is unsure about a music structure decision (e.g. whether to repeat a section or move on) and the musician needs to look ahead in the music, then a display must be capable of showing at least three lines (or "systems") of music at once: the current line plus the two alternative destinations. This provides a way to structure page turning and page layout on a dynamic music display.

The Vision of HCMP

To develop a practice of HCMP and to build upon these initial investigations, we need to imagine how humans and computers will interact, what sorts of communication will take place, and what sorts of processing and machine intelligence will be needed. We need a research agenda. To guide this imagining process, we should look at the practice of music performance *without* computers. From this, we will construct a set of *predictions* that anticipate characteristics and functions of future HCMP systems. These predictions will serve to guide future investigations and pose challenges for research and development. We can divide HCMP into two main activities: music preparation and music performance.

Music Preparation

An assumption in HCMP is that music is well-structured: There are agreed-upon melodies, chord progressions, bass lines, and musical structure that must be communicated to all performers. If the music performance is always the same, this is trivial, but our assumption is that the structure may change even during the performance. What happens when the vocalist decides to sing the verse again or the bandleader directs the band to skip the drum solo? This relates to the descriptions of static and dynamic scores.

We can think of static scores as sequential computer programs. The score is "executed" by performing one measure after the next. Musical repeats, jumps, and optional endings are program control constructs (loops, gotos, and conditionals). The dynamic score is then a trace of the execution of this program. With this analogy, one can imagine the preparation of a computer performer to be a kind of programming: "When you reach measure 17, if this is the second repeat, then if there is no human bass player, then play this audio." A conventional programming language is certainly not the right way to express these "programs," but it is clear that we will need something more than a conventional audio recorder and editor. Designing interfaces that are both intuitive and expressive for "programming" performances is an important problem. *Predictions: HCMP systems will make the static/dynamic score relationship more explicit. Terminology for specifying the location in a dynamic score in terms of the static score will be formalized. Score location will be indicated not only in terms of measure numbers but also in terms of the static score structure. Techniques for displaying and directing dynamic score location will form a necessary part of the communication between human and computer performers.*

"Scores" in popular music performance can range from complete and detailed common music notation (as in "classical" works) to highly abstract descriptions such as lyrics or lists of sections. Other music representations are also common: drummers often need just the music structure (how many measures in each section) without actual instructions on what to play, and keyboard, bass, and guitar often read from "chord charts" that give chord symbols rather than specific pitches. *Prediction: HCMP systems will work with multiple music representations.*

Computer-generated music can be based on audio (with time stretching for synchronization), MIDI sequences, or computer composition and improvisation from specified chord progressions. For many musical genres, automatic generation of parts is feasible, as illustrated by programs such as Band-in-a-Box (Gannon 2004). However, there are seemingly infinite varieties of styles and techniques, so there is plenty of room for research in this area. An interesting problem is not just to, say, create a bass line in a given style, but to give the user control over different parameters or to allow the user to say "I want a bass line like the one in song X," where of course song X has a different key, tempo, and chord progression. This is a kind of musical analogy problem (Hofstadter, 1996): bass line *a* is to music structure *b* as bass line *c* is to music structure *d*. Given *b, c,* and *d,* solve for *a*. Many users will not have the skill, time, or inclination to play the parts themselves or compose the parts note-by-note, so the ability to generate parts automatically is an essential feature. *Prediction: HCMP systems will rely on stylistic generation of music according to lead sheets in addition to pre-recorded audio and sequenced MIDI data.*

Music notation offers a direct visual and spatial reference to the otherwise ephemeral music performance. As discussed earlier, we envision capturing music notation by camera or scanner (Lobb, Bell, and Bainbridge 2005) as well as using computer-readable notation. For unstructured images, one would like to convert the notation into a machine-readable form, but like OCR, optical music recognition (OMR) is far from perfect, especially for hand written (or scrawled) lead sheets. Furthermore, some musicians work from lyrics and chord symbols rather than common music notation. It seems essential to develop methods to annotate music images with structural information. In most cases, this annotation of music notation will be the mechanism by which the static score is described and communicated to the computer. *Prediction: HCMP systems will extend music notation to specify music structure.*

One characteristic of popular music performance addressed by HCMP is the preparation of "scores" before the performance. Unlike most classical music where the score is carefully prepared by the composer and publisher, popular music is more likely to be arranged and structured by the performing musicians. Decisions to alter the introduction, where and how to end, and whether to repeat sections are common. *Prediction: HCMP systems will provide interfaces for specifying arrangements and performance plans.*

Having discussed audio, MIDI, and various forms of music notation, it should be obvious that an important function of HCMP systems will be to provide abstractions of music structure and to allow users to integrate and coordinate multiple representations of music. *Prediction: A primary function of HCMP systems will be to coordinate multiple media both in preparation for and during live performance.*

Music Performance

Once parts are prepared, we need to perform them! The main issues have to do with musical synchronization and communication. Indeed, the primary reason that there is no common practice of HCMP today is the difficulty of getting artificial performers to synchronize to live music. There were early attempts to achieve HCMP using tape recorders and other technologies, but these were mostly abandoned. Many street musicians and solo acts today use a very limited form of HCMP in which the performer simply switches on a pre-recorded "backup band" and plays or sings along. This same idea is seen in Karaoke and many TV and theater productions. B-Keeper is a recent system that uses live audio for beat-based synchronization (Robertson and Plumbley 2007). We want to envision how performers and larger groups could function if many of their present limitations were removed.

When musicians perform together, they synchronize at several levels of a time hierarchy. At the lowest level is the beat or pulse of the music. Unfortunately, fast and accurate automatic detection of beats is not a solved problem. *Prediction: HCMP systems will use a variety of beat detection systems and integrate information from multiple sources in order to achieve the necessary accuracy and reliability to support computer-based performers.*

Another level of time synchronization is the *measure* (or *bar*). Typically a group of 2 or 4 beats, measures organize music into chunks. In rock, measures are indicated by the familiar snare drum accents on beats 2 and 4 and chord changes on or slightly before beat 1. Measures are important in music synchronization because sections are aligned with respect to measures. A musician would never say "let's go to section B on the 3rd beat of measure 8." *Prediction: HCMP systems will track measure boundaries. As with beats, multiple sensors and modalities will be used to overcome this difficult machine listening problem.*

Finally, music is organized into sections consisting of groups of measures. These sections are typical units of arrangement, such as introductions, choruses, and verses. When a performance plan is changed during the performance, it is usually accomplished by communicating, in effect, "Let's play section B now (or next)." In the case of "now," the section begins at a measure boundary. In the case of "next," the new section begins at the end of the current section. Without these higher-level temporal structures and conventions, synchronization and cues would need to be accurate to the nearest beat, perhaps just a few hundred milliseconds rather then the 1 to 10 seconds afforded by higher level structures. *Prediction: HCMP systems will be "aware" of measures and higher-level sectional boundaries in order to synchronize to human players. As with measures, multiple sensors and modalities will be used to overcome the machine listening problem of identifying musical sections.*

Two Examples

It is useful to describe sessions with imagined HCMP systems in order to grasp how the overall system might function. I will describe two examples. The first is a rehearsal and private practice with a conventional orchestra. The second is an informal jam session.

The orchestra example will focus on music display and practice as opposed to computer performance. In fact, it falls outside the assumptions of popular music, improvisation, and steady tempo, but it is good to show that these restrictions are not always needed. Imagine that ordinary music on paper is available before the first rehearsal. Using a camera, each page is captured as a digital image and wirelessly transferred to a digital tablet. OCR and OMR do some preliminary analysis of the images to identify titles, rehearsal markings, and staff and measure locations. The captured information is identified using fonts or colors so that the musician can visually confirm where the automatic recognition is correct and intervene where recognition was in error or missing.

Printed music is often on large pages arranged side-by-side on a music stand to minimize page turning. While a folding digital display might be able to reproduce this arrangement, we will assume a smaller display that necessitates more frequent page turning or scrolling. (Bell, *et al.* 2005) The semi-automated staff recognition divides the music into sub-page units that can be displayed in sequence. These units of music are automatically arranged from top to bottom on the display. When the player reaches the bottom, the next unit of music overwrites the top of the display, allowing the musician to read ahead.

Because of repetitions in the music, the display is not strictly sequential through the pages. Using either automated or manual markup techniques, repeats and other markings can be identified, and the tablet can show the music in the proper *dynamic* order.

After practicing parts at home, the musician takes the tablet to the first orchestra rehearsal. There, the tablet offers a directory of pieces and an index into rehearsal markings and measure numbers so that the musician can quickly jump to any location requested by the conductor. The tablet records audio from the entire rehearsal and tags the audio with score locations based on whatever music is being displayed at that moment. In the rehearsal, music can be advanced by eye tracking, foot pedals, or other sensors. One viable and simple method is a ring on the first finger that can be pressed with the thumb to signal a page turn. Once music has been rehearsed, the tablet might take a more active role in page turning by matching the live music to recordings from previous rehearsals (Dannenberg and Raphael 2006).

Back at home, as the musician resumes practice, recordings from rehearsals can be selected, allowing the musician to play along with the sound of the orchestra. Ideally, one might want to remove the part to be practiced from recordings. There are a couple of techniques that might at least suppress the unwanted sound (Han and Raphael 2007, Smaragdis and Mysore 2009). Another practice aid is the ability to speed up or slow down the rehearsal audio using time stretching techniques. This is in fact already available in the SmartMusic (MakeMusic 2010) commercial practice system, but SmartMusic does not integrate the capture of scores and rehearsal audio.

If music tablets communicate, and if their score representations are comparable, then it will be possible for the conductor to direct everyone's attention to a particular location (in practice, much rehearsal time is currently spent directing musicians to particular locations in the score … "please look at the third beat of the fourth measure after letter G … no, the fourth measure … yes, the third beat … yes, where you have an F-sharp …"). Page turning could be made more reliable and automatic by sharing location and confidence estimates among dozens of tablets.

The next example is a jam session. Imagine that some friends want to play some songs they have played together before, but a bass player is not available. To prepare for the session, the leader finds music on the Web consisting of MIDI files or some commercial formats such as Guitar Pro (Arobas 2010) or Band-in-a-Box (Gannon 2004). Using an HCMP system, the music is imported and automatically converted into a "lead sheet" representation, which has the music structure and chord names. This may require automatic analysis to derive chords and music structure. The user may then reorganize the music into a performance plan such as "4-bar intro, verse, chorus, verse, ending."

Rather than laboriously preparing each song, the band leader might download ready-to-use bass parts for the songs from the Web. These might be posted to sharing sites similar to those currently storing MIDI, guitar tablature, lyrics, and other music data. Or, there might be commercial sites offering virtual musicians and song data for them to play, just as one can now buy ready-to-use clip-art and background music for multimedia productions.

At the jam session, the group selects a song to play (informing the computer), and the leader counts off the beginning. The computer joins the performance. The entrance could be synchronized by many mechanisms that include foot pedals, gestures, speech recognition to "hear" the count-off, etc. Once the band is playing, the bass stays in time by synchronizing to beats. Again, there are many possible ways to detect beats, including foot tapping, audio analysis, and gestures detected by vision, inertial sensors, or other techniques. (No current systems can do this well.)

As the band rehearses, there may be directions for the computer bass player to adjust the sound, the volume, the style, etc. This will require an interface where musical style can be manipulated. In addition, the computer must generate a bass line from the chord representation. This is a computer composition or improvisation task that could be accomplished off line or in real time.

During the performance, the band may decide to play the chorus an extra time (for example). Human performers might signal this by gesture or by shouting "chorus" a measure before the repetition should begin. It seems unlikely that the computer will be able to understand natural human gestures like this, but there are many ways to communicate the information, such as touching the beginning of the chorus on a tablet-based music display (in which case the computer would understand that, when the current section finishes, it should go back to the pointed to location). A gesture-based interface might accomplish the same task.

Future Work

Our work to date has focused on identifying the potential applications and functions of HCMP systems by building experimental systems, using them, and speculating how computers might be used in future systems. This paper has described three prototype systems that illustrate some of the musical potential of this work. Based on these prototypes, we have identified a number of interesting problems to pursue in future work. We have made specific predictions about the major characteristics of future systems, establishing a research agenda which we now summarize.

An important area for research is preparation of musical scores. While computer-based music notation editors exist, they work at the note and measure level, whereas HCMP systems should allow users to alter existing music in terms of sections. (E.g. "play the chorus twice.") Another limitation of existing editors for music is the inability to deal with multiple representations (lyrics, notes, chords, etc.) and multiple media (notation, audio, MIDI). While specialized editors are still important, we need a way to integrate and coordinate all representations used in the music performance.

A second area for research is synchronization. Computers must maintain a representation of at least three levels of the time hierarchy: beats, measures, and sections. A variety of techniques and modalities can be used. Since even humans use both audible and visual cues, it seems that HCMP systems will need to integrate multiple sources of timing information and exchange information between different levels of the timing hierarchy in order to provide reliable synchronization.

More work is also needed to design systems that clearly express the relationship between static scores and their unfolding into a linear performance. Problems include naming (how to refer to a specific location in the dynamic score), communicating intentions before and during music performances, and adapting either recorded or generated music to dynamic changes to the score.

The actual sound generated by the computer is important. Whether the sound is from recordings or is generated on-the-fly, users will need support to prepare and control this sound. This includes problems of synthesis and sound diffusion. Perhaps the most important musical issue is *control*. How will humans "ask" the computer player to play in different styles, and how will style be represented and realized in a controllable fashion?

Conclusions

Human Computer Music Performance is at present more of a dream than reality. This paper offers a set of research problems based on experience from a few early HCMP systems and thinking hard about how current practice in live music performance can be extended with state-of-the-art computation. Due to the deep musical knowledge and experience needed for musical interaction in popular forms of music, it seems that the solutions lie in careful design that balances human-computer interaction techniques with machine intelligence. The former reduces the need to automate musical intelligence completely while the latter reduces the burden of direct human control and intervention. It will be very exciting if HCMP reaches its potential to impact thousands or even millions of musicians. Our prototypes illustrate that HCMP does not require any technical breakthroughs to be practiced in a simple form now, but the deeper issues of music generation and music understanding will likely mature over the next decade. Once it is established and widely available to artists, we believe HCMP will inspire and enable new concepts and genres in music that cannot yet be imagined.

Acknowledgements

Thanks to Ryan Calorus, who implemented our experimental music display, and Nicolas Gold for valuable discussions and pointing out the importance of (re)arrangement by music sections. Dawen Liang and Gus Xia assisted with the most recent performance system. 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 for our use. Current work is supported by the National Science Foundation under Grant No. 0855958.

References

Arobas Music, Inc. 2010. Guitar Pro 6 (software).

Bell, T.; Blizzard, D.; Green, R.; and Bainbridge, D. 2005. Design of a digital music stand. *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR06)*: 430-433.

Connick, H. Jr. 2002. System and method for coordinating music display among players in an orchestra. US Patent #6348648.

Dannenberg, R. 2011. "A virtual orchestra for human-computer music performance." In review.

Dannenberg, R. and Raphael, C. 2006. Music score alignment and computer accompaniment. *CACM* **49**(8): 38-43.

Gannon, P. 2004. Band-in-a-Box (software), PG Music.

Han, Y. and Raphael, C. 2007. Desoloing monaural audio using mixture models. *Proceedings of the 8th International Society of Music Information Retrieval Conference (ISMIR 2007).*

Hofstadter, D. 1996. *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York: Basic Books.

Lobb, R.; Bell, T.; and Bainbridge, D. 2005. Fast capture of sheet music for an agile digital music library. *ISMIR 2005 6th International Conference on Music Information Retrieval Proceedings*: 145-152.

MakeMusic, Inc. 2010. SmartMusic (software).

Robertson, A. and Plumbley, M. 2007. B-keeper: A beat-tracker for live performance. *Proceedings of the International Conference on New Interfaces for musical expression (NIME)*: 234–237.

Schnell, N.; Peters, G.; Lemouton, S.; Manoury, P.; and Rodet, X. 2000. Synthesizing a choir in real-time using pitch-synchronous overlap add (PSOLA). *Proceedings of the International Computer Music Conference.*

Smaragdis, P. and G. Mysore. 2009. Separation by "humming": User-guided sound extraction from monophonic mixtures. *Proceedings of IEEE Workshop on Applications Signal Processing to Audio and Acoustics.*