

THE MACHINE'S GOT RHYTHM

Computers are learning to understand music and join the band

BY JULIE J. REHMEYER

Christopher Raphael begins the third movement of a Mozart oboe quartet. As his oboe sounds its second note, his three fellow musicians come in right on cue. Later, he slows down and embellishes with a trill, and the other players stay right with him. His accompanists don't complain or tire when he practices a passage over and over. And when he's done, he switches them off.

After all, his fellow musicians exist only as a recording. A software package, written by Raphael, controls their tempo and makes them respond to the soloist's cues.

Until recently, computers have had little insight into music. They've merely recorded it, stored it, and offered tools that people can use to produce or manipulate it. But now, researchers are teaching computers to recognize the basic musical elements: beat, rhythm, melody, harmony, tempo, and more. Computers with those skills are becoming musical collaborators.

"Technology is changing our sense of what music can be," Raphael says. "The effect is profound."

LEARNING TO LISTEN With training, people can listen to a piece of music and write down the score with few mistakes. Teaching a computer to perform the same task, though, has proved remarkably difficult.

Raphael, an informatics researcher at the University of Indiana in Bloomington, compares the problem to speech recognition. "There's been a veritable army of people who've worked on speech recognition for several decades, and [the problem] still remains open," he says. "Any time you deal with real data, there is a huge amount of variation that you have to understand."

Researchers have succeeded in programming computers to transcribe limited kinds of music. For example, software can reliably identify the notes of a single melodic line played by one instrument in isolation.

The programs analyze the wavelengths of the sound. Hitting the A below middle C on a piano, for example, produces an audio wave at 220 Hertz. But it also produces weaker waves, known as

overtones, at 440 Hz, 660 Hz, 880 Hz, and so on. The relative strengths of the overtones differ slightly for each instrument, which is why a piano doesn't sound like a violin. Nevertheless, the characteristic pattern of an A is similar enough across instruments that a computer can recognize it.

When several notes play simultaneously, however, as in a chord from one instrument or music from an ensemble, the audio waves from the different notes mix in ways that are hard to untangle. Echoes, noise, and imperfect recordings muddy the patterns even more.

But researchers are making progress. Every year, various transcription programs go head-to-head in a competition called MIREX (Music Information Retrieval Exchange). The researchers

set their programs loose on the same pieces of music and then compare results. This September, when the competition takes place in Vienna, it will for the first time include full transcriptions of polyphonic music, in which multiple notes are playing at the same time.

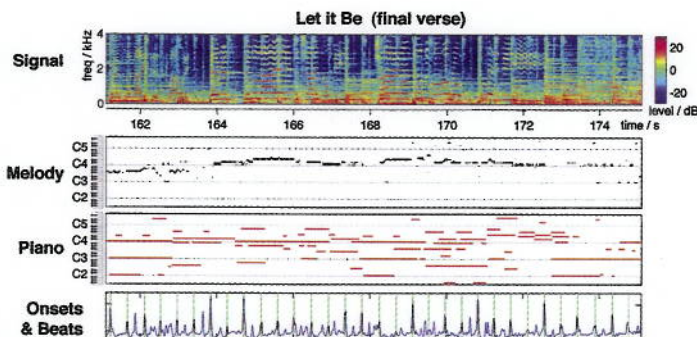
Most systems slice the sound into brief segments and look for a pattern that they can recognize as a given note. After identifying this note, the programs pull its primary frequency and associated overtones out of the sound wave. Then the software repeats the process,

picking out other notes in the remaining audio signal until it has accounted for the entire sound.

The results, however, aren't exact. The pattern of a particular note may be obscured by other notes that are playing at the same time. Furthermore, without information on the characteristics of the instrument producing the sound or the acoustics of the room in which it was recorded, the programmed patterns of overtones don't accurately correspond to the actual notes in the music.

As a result, when the program pulls an imperfectly modeled note out of the mix, it distorts the remaining sound, making it harder to identify the remaining notes. The more notes that are playing at once, the more those distortions pile up.

SELF-TEACHING MACHINES Music-information researchers are taking advantage of the experiences of their colleagues who study speech recognition. After some early advances in the 1970s, further improvements in speech recognition became increasingly difficult. "To take it to the next level," says Daniel Ellis of Colum-



COMPUTING A SOUND — At top, the spectrogram for the final verse of "Let It Be" shows how the sound intensity changes over time at each frequency. A computer program detects the beats, the melody, and a piano-roll version of the full score, with horizontal stripes indicating the activation of particular notes.



bia University, “you had to do 10 times as much work each time.”

By the time Ellis started working on speech recognition in 1996, researchers were trying a new approach. “To some extent, they gave up on trying to understand what speech does,” Ellis says. “Instead, they collected a bunch of different examples and used statistical techniques” to identify the patterns that underlie speech.

Ellis continued that strategy when he eventually shifted his focus to the analysis of music. He built a program that uses machine-learning techniques to transcribe polyphonic piano music.

He started with a program that had no information about how music works. He then fed into his computer 92 recordings of piano music and their scores. Each recording and score had been broken into 100-millisecond bits so that the computer program could associate the sounds with the written notes. Within those selections, the computer would receive an A note, for example, in the varying contexts in which it occurred in the music. The software could then search out the statistical similarities among all the provided examples of A.

In the process, the system indirectly figured out rules of music. For example, it found that an A is often played simultaneously with an E but seldom with an A-sharp, even though the researchers themselves never programmed in that information. Ellis says that his program can take advantage of that subtle pattern and many others, including some that people may not be aware of.

When presented with a novel recording, the program labels as an A any note that shows enough statistical similarity to the As in the training sequence. In a special issue of *EURASIP Journal on Advances in Signal Processing*, an online journal, Ellis reports that his system accurately identified the notes playing in 68 percent of the novel 100-millisecond snippets that it was given. Ellis expects that when his program has analyzed more examples—ideally, many thousands more—its detection rate will improve.

He notes that the next-best system, developed by Anssi Klapuri of the Tampere University of Technology in Finland, scored only 47 percent on the test snippets. It’s a traditional program that incorporates expert knowledge of music rather than machine learning.

Ellis is quick to point out, however, that this comparison isn’t quite fair. Klapuri’s system can recognize many kinds of music, not just piano music, so comparing the two on piano music alone gave Ellis’ system an artificial advantage.

Ellis plans to enter his program in the September 2007 MIREX competition to see how it does head-to-head against more-traditional programs.

Ellis has also used the self-teaching technique to identify melodies in complex pieces of music, picking out the portion that a person might sing. After spending just a few months to develop such a system, he entered it in last year’s MIREX competition and came in third out of 10 entries, with an accuracy of 61 percent. In many cases, he says, the transcribed melodies were recognizable, despite the errors.

The top performer in that competition was a more fully developed program that took a traditional approach. Devised by Karin Dressler of the Fraunhofer Institute for Digital Media Technology in Ilmenau, Germany, that program had a 71 percent accuracy rate. The results of the melody competition will appear in an upcoming issue of *IEEE Transactions on Audio, Speech and Language Processing*.

Ellis says that combining machine-learning strategies with expert knowledge of music and acoustics will ultimately offer the best performance.

FOLLOWING THE MUSIC Even as researchers continue to refine transcription methods, the work is spinning off remarkably useful tools. One advance has turned out to be especially handy: Computers can line up a score with a recording of its performance.

This seemingly trivial capability has many applications. Some of the simplest are programs that display supertitles at the opera at just the right moment or that automatically turn the page for musicians.

Score alignment also opens the door to programs that can correct off-kilter notes going into a microphone before they emerge from loudspeakers—a development that could transform the listener’s experience at children’s recitals everywhere.

Alignment software analyzes a spectrogram, which shows how the energy of sound waves changes over time across all frequencies. In most popular music, the strong drum rhythms that mark out the time appear on the spectrogram as vertical lines, which make it easy for the computer to keep track of where it is in the score. Another approach that some programs use is to recognize repeating harmonic patterns that occur in many pieces of music.

Where drumbeats or repeating harmonic patterns aren’t apparent, the researchers have the computer identify the melody or employ other techniques developed for transcription. Having the score as a guide makes the task far easier than transcribing the notes from scratch.

Score-alignment programs could be used after a musician records a piece of music to do the kind of fine-tuning that’s now performed painstakingly by recording studios, fixing such problems as notes that are slightly off pitch or come in late. “It’ll be kind of like a spell-check for music,” says Roger Dannenberg, a computer scientist at Carnegie Mellon University in Pittsburgh who is developing the technology.

The process would make it far easier for amateurs to improve their recordings after performance in the way that professional recording studios now do. “I see what I’m doing as democratizing music-making,” Dannenberg says.

COMPUTER AS MUSICIAN Score-alignment technology opened the door for Raphael to develop his computerized-accompaniment program. Mimi Zweig, a professor of music at the University of Indiana, is using the system with her violin students to give them a taste of what it’s like to have 100 musicians following their every pause or trill.

Zweig is impressed with the responsiveness of the system. “After a long cadenza or a phrase where you want to take time, it’s right with you,” she says. “It’s even better than an orchestra in some ways.”

Raphael says that the soloist’s freedom while using his system makes it a valuable learning tool. Few students ever experience having an orchestra accompany them. Raphael says, “It’s a fundamental hole in their musical education. [Playing with an orchestra] is how people develop their ideas about musical interpretation and grow as musicians.”

The first component of Raphael’s program examines the sound waves produced by the soloist and lines up the performance with the score. But that’s not enough, because if the program waits until the soloist plays a note before it comes in with the accompaniment, it will always be late. So, the program predicts what the soloist will do next, using information about the performance from which the accompaniment was derived and the performer’s speed in the immediately preceding notes as well as knowledge gained from earlier practice sessions. The program then slows down or speeds up the recording without altering the pitch.

