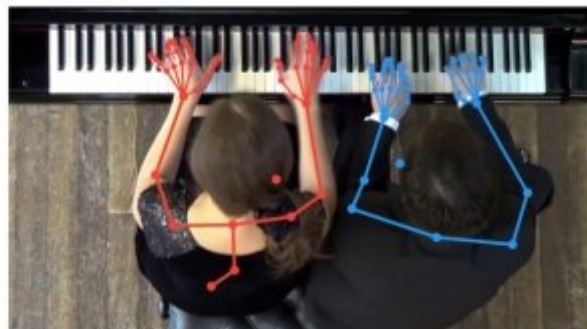


# Identifying a melody by studying a musician's body language

June 26 2020, by Kim Martineau



Researchers use skeletal keypoint data to match the movements of musicians with the tempo of their part, allowing listeners to isolate similar-sounding instruments. Credit: Massachusetts Institute of Technology

We listen to music with our ears, but also our eyes, watching with

appreciation as the pianist's fingers fly over the keys and the violinist's bow rocks across the ridge of strings. When the ear fails to tell two instruments apart, the eye often pitches in by matching each musician's movements to the beat of each part.

A [new artificial intelligence tool](#) developed by the MIT-IBM Watson AI Lab leverages the virtual eyes and ears of a computer to separate similar sounds that are tricky even for humans to differentiate. The tool improves on earlier iterations by matching the movements of individual musicians, via their skeletal keypoints, to the tempo of individual parts, allowing listeners to isolate a single flute or violin among multiple flutes or violins.

Potential applications for the work range from sound mixing, and turning up the volume of an instrument in a recording, to reducing the confusion that leads people to talk over one another on a video-conference calls. The work will be presented at the virtual Computer Vision Pattern Recognition conference this month.

"Body keypoints provide powerful structural information," says the study's lead author, Chuang Gan, an IBM researcher at the lab. "We use that here to improve the AI's ability to listen and separate sound."

In this project, and in others like it, the researchers have capitalized on synchronized audio-video tracks to recreate the way that humans learn. An AI system that learns through multiple sense modalities may be able to learn faster, with fewer data, and without humans having to add pesky labels to each real-world representation. "We learn from all of our senses," says Antonio Torralba, an MIT professor and co-senior author of the study. "Multi-sensory processing is the precursor to embodied intelligence and AI systems that can perform more complicated tasks."

The current tool, which uses body gestures to separate sounds, builds on

earlier work that harnessed motion cues in sequences of images. Its earliest incarnation, PixelPlayer, let you click on an instrument in a concert video to make it louder or softer. An update to PixelPlayer allowed you to distinguish between two violins in a duet by matching each musician's movements with the tempo of their part. This newest version adds keypoint data, favored by sports analysts to track athlete performance, to extract finer grained motion data to tell nearly identical sounds apart.

The work highlights the importance of visual cues in training computers to have a better ear, and using sound cues to give them sharper eyes. Just as the current study uses musician pose information to isolate similar-sounding instruments, previous work has leveraged sounds to isolate similar-looking animals and objects.

Torralba and his colleagues have shown that deep learning models trained on paired audio-video data can learn to recognize [natural sounds](#) like birds singing or waves crashing. They can also pinpoint the geographic coordinates of a moving car from the sound of its engine and tires rolling toward, or away from, a microphone.

The latter study suggests that sound-tracking tools might be a useful addition in self-driving cars, complementing their cameras in poor driving conditions. "Sound trackers could be especially helpful at night, or in bad weather, by helping to flag cars that might otherwise be missed," says Hang Zhao, Ph.D. '19, who contributed to both the motion and sound-tracking studies.

**More information:** Music Gesture for Visual Sound Separation: arXiv:2004.09476 [cs.CV] [arxiv.org/abs/2004.09476](https://arxiv.org/abs/2004.09476)

*This story is republished courtesy of MIT News*

([web.mit.edu/newsoffice/](http://web.mit.edu/newsoffice/)), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Identifying a melody by studying a musician's body language (2020, June 26) retrieved 9 April 2024 from <https://techxplore.com/news/2020-06-melody-musician-body-language.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.