

# Computers sing to a better tune

May 6 2011

---

Would you dance to a synthetic song? A new approach to making computer-generated vocals more "human" is reported this month in the *International Journal of Knowledge Engineering and Soft Data Paradigms*.

Music producers have for decades had electronics tricks at their disposal for improving a recorded vocal performance. They can add a little reverb or echo to bolster a weak rendition, use effects such as phasing and delay to add color to the vocal, fix duff notes with auto-tuning or even reprogram a whole melody line in software. In recent years, [voice synthesis](#) for converting text to spoken word has improved considerably but combining that technology with auto-tuning capability allows computers to "sing".

Software, such as Vocaloid, can successfully create lead vocals and harmony parts from an input of lyrics and musical score. Careful tweaking of the "frequency curve" can make the vocals sound almost natural by adding tremolo, vibrato and note overshoot.

Researchers in the Graduate School of Engineering, at The University of Tokyo, point out that it is the tweaking of the frequency curve that is critical to success but this process is labor intensive and prone to human error that means a vocal rendition always retains artifacts of the synthetic process used to produce it. Other researchers have developed tools such as VocaListener, which can tune frequency curves and make some features of singing [voice](#) like vibrato. However, this system requires an original version of the vocal sung by a human from which to

work. The SingBySpeaking system "sings" text input, but uses only one frequency curve and so is insufficient for a realistic synthetic vocal with all the nuance of a human performance.

Now, Akio Watanabe and Hitoshi Iba have turned to evolution to help them devise a novel algorithm that compares the frequency curves from real human performances and uses them to home in on a more realistic curve to apply to the synthetic song. The team has simplified the optimization process for creating vocal frequency curves and have developed a frequency model that can emulate human expression in a synthetic vocal.

There are four steps to the evolutionary process for creating a realistic frequency curve, explain Iba and colleagues:

First, production of the first generation involves making eight individual curves with random parameters and feeding them into Vocaloid. The second step is for the music producer to listen to the effect of each curve on their synthetic vocal and to move slider bars in the software interface to reflect how well each curve works. In the third stage, the best curves are used as the "parents" to create a new generation of curves. Finally, the second generation curves undergo crossover and random mutation and the process repeated from step 2. Eventually, the fittest frequency curves will emerge that endow the synthetic vocal with the most realistic characteristics of human [singing](#).

For anyone who is bored with the so-called real-life characters that present themselves to TV "talent" shows, an optimized frequency curve and a synthetic vocal could be the new sensation they are looking for, but without the baggage of bad teeth, terrible hair extensions and fictionalized family tragedies.

**More information:** "Creating singing vocal expressions by means of

interactive evolutionary computation" in Int. J. Knowledge Engineering and Soft Data Paradigms, 2011, 3, 40-56

Provided by Inderscience Publishers

Citation: Computers sing to a better tune (2011, May 6) retrieved 26 April 2024 from <https://phys.org/news/2011-05-tune.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.