

# Disney Research creates method enabling dialogue replacement for automated video redubbing

April 20 2015

---



A badly dubbed foreign film makes a viewer yearn for subtitles; even subtle discrepancies between words spoken and facial motion are easy to detect. That's less likely with a method developed by Disney Research that analyzes an actor's speech motions to literally put new words in his mouth.

The researchers found that the facial movements an actor makes when saying "clean swatches," for instance, are the same as those for such phrases as "likes swats," "then swine," or "need no pots."

Sarah Taylor and her colleagues at Disney Research Pittsburgh and the University of East Anglia, report that this approach, based on "dynamic visemes," or [facial movements](#) associated with [speech](#) sounds, produces far more alternative word sequences than approaches that use

conventional visemes, which are static lip shapes associated with sounds.

They will present their findings April 23 at ICASSP 2015, the IEEE International Conference on Acoustics, Speech and Signal Processing in Brisbane, Australia.

"This work highlights the extreme level of ambiguity in visual-only speech recognition," Taylor said. But whereas a lip reader battles against ambiguity, using context to figure out the most likely words that were spoken, the Disney team exploited that ambiguity to find alternative words.

"Dynamic visemes are a more accurate model of visual speech articulation than conventional visemes and can generate visually plausible phonetic sequences with far greater linguistic diversity," she added.

The relationship between what a person sees and what a person hears is complex, Taylor noted. In the "clean swatches" example, for instance, the alternative word sequences can vary in the number of syllables yet remain visually consistent with the video.

Speech redubbing, such as translating movies, television shows and video games into another language, or removing offensive language from a TV show, typically involves meticulous scripting to select words that match lip motions and re-recording by a skilled actor. Automatic speech redubbing, as explored in this study, is a largely unexplored area of research.

With conventional static visemes, a lip shape is assumed to represent a small number of different sounds and the mapping of those units to distinctive sounds is incomplete, which the researchers found limited the number of alternative word sequences that could be automatically

generated.

Dynamic visemes represents the set of speech-related lip motions and their mapping to sequences of spoken sounds. The researchers exploit this more general mapping to identify a large number of different word sequences for a given facial movement.

"The method using dynamic visemes produces many more plausible alternative word sequences that are perceivably better than those produced using a static viseme approaches," Taylor said.

Provided by Disney Research

Citation: Disney Research creates method enabling dialogue replacement for automated video redubbing (2015, April 20) retrieved 23 April 2024 from <https://medicalxpress.com/news/2015-04-disney-method-enabling-dialogue-automated.html>

<p>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.</p>
--