Audio Morphing

Tony Ezzat

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: In this work, we tackle the problem of morphing between different audio sequences. The system should take as input 2 audio sequences, and produce as output intermediate audio sequences that represent natural exemplars lying between the 2 input sequences.

Motivation: Audio morphing might have important applications in speech recognition, speech synthesis, music synthesis, and other applications where large corpora are recorded and there is a strong need to interpolate between the exemplars in the corpora to produce new exemplars.

Previous Work: There has been a spate of recent work on *voice conversion* [4, 5, 2], where a reference speaker speech sample is warped to match the statistical properties of a target speaker. Most authors resort to mixed time- and frequency- domain methods to alter pitch, duration, and spectral features.

Audio morphing [3] is closest in spirit to the goal of this work. The author used dynamic time-warping to time-align two speech samples, cross-faded the respective smoothed spectrograms, and warped a pitch residual to morph between two sounds.

Approach: In this work, we aim to explore and develop a purely time-domain method of audio morphing, motivated by the success of methods such as TD-PSOLA [1] for warping audio sequences.

Research Support: ONR, MURI, Darpa, NSF, Air Force, DaimlerChrysler, Honda, Siemens, Kodak, ATT, ATR, Compaq, CRIEPI

References:

[1] E. Moulines and F. Charpentier Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones *Speech Communication* 9 (1990) pg 453–467

[2] H. Valbret, E. Moulines, and J.P. Tubach Voice Transformation using PSOLA Technique *Speech Communication* 11 (1992) pg 175–187

[3] M. Slaney, M. Covell and B. Lassiter Automatic Audio Morphing *Proc. ICASSP*, Atlanta Georgia, May 7-10 1996

[4] A. Kain and M. Macon Spectral Voice Conversion for Text-to-Speech Synthesis *Proc. ICASSP*, May 1998, pg 285–288

[5] Y. Stylianou O Cappe and E. Moulines Statistical Methods for Voice Quality Transformation *Proc. Eurospeech*, September 1995, Madrid Spain, pg 447-450