wakayama univ.

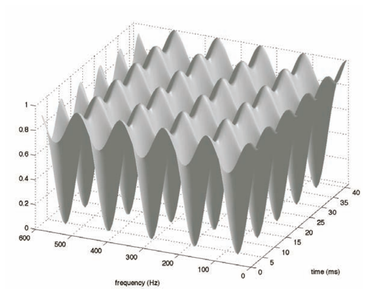
国立大学法人
和歌山大学

CrestMuse Symposium 2008
http://www.crestmuse.jp/sympo2008/

STRAIGHT, TANDEM-STRAIGHT and beyond

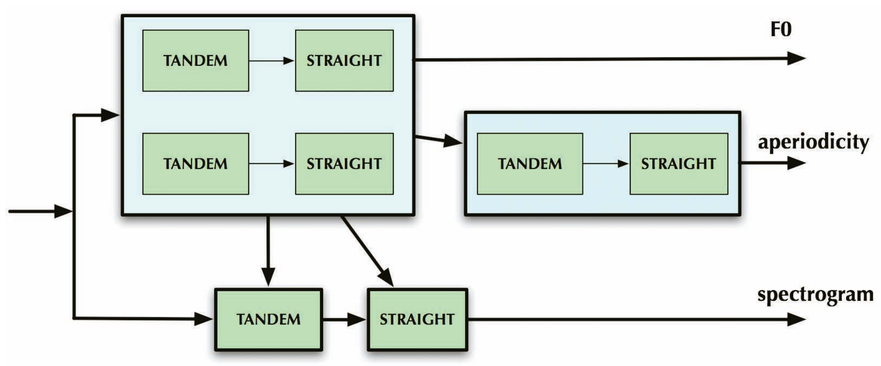
Hideki Kawahara, Masanori Morise, Toru Takahashi, Ryuichi Nisimura
Hideki Banno and Toshio Irino
(Wakayama Univ., Kwansei Gakuin Univ., Kyoto Univ. and Meijo Univ.)

Overview/Summary



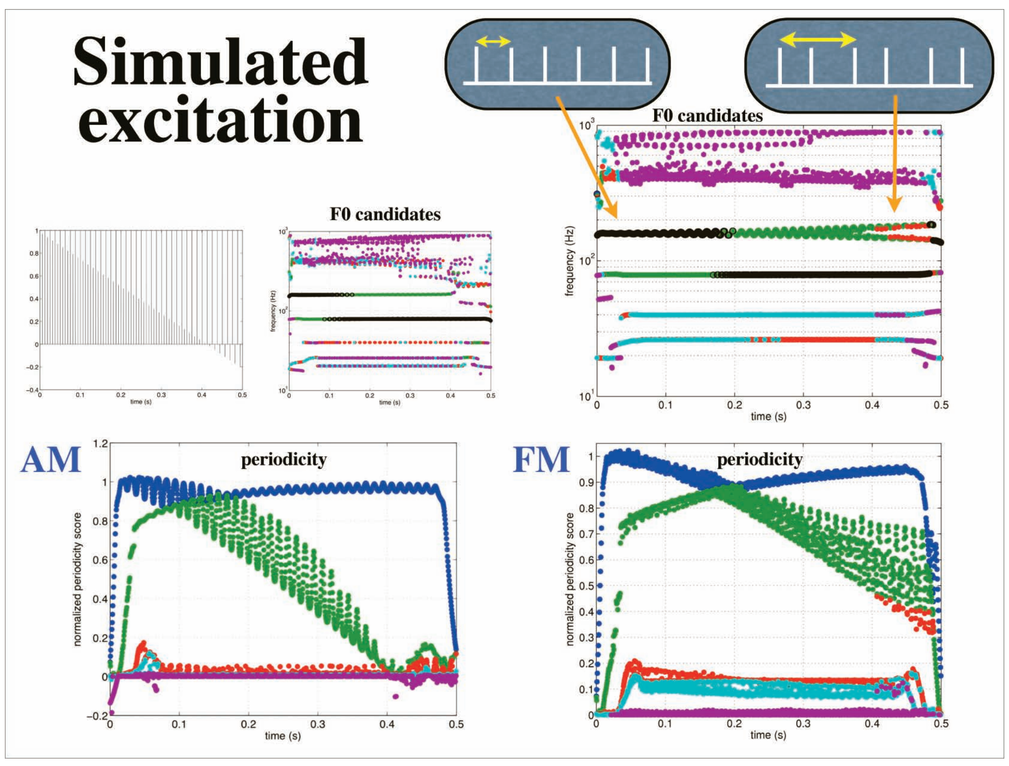
STRAIGHT, a speech analysis, modification synthesis system, is an extension of the classical channel VOCODER that exploits the advantages of progress in information processing technologies and a new conceptualization of the role of repetitive structures in speech sounds. STRAIGHT was designed to provide representation consistent with our auditory perception which decomposes input sounds in terms of excitation (source) and resonant (filter) characteristics. This architecture and sophisticated implementation made it the most flexible and high-quality speech manipulation system to date. However, underlying algorithms were not well formulated mathematically sound manner.

TANDEM, a method to extract temporally stable power spectral representation and consistent sampling theory, a new mathematical foundation of sampling made possible to reformulate STRAIGHT completely based on sound foundation. It also simplified codes and enhanced execution speed drastically.



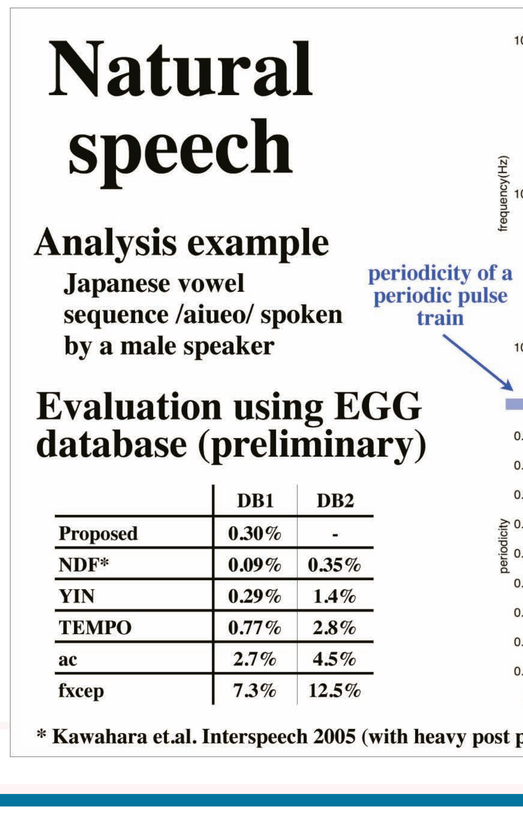
TANDEM-STRAIGHT architecture

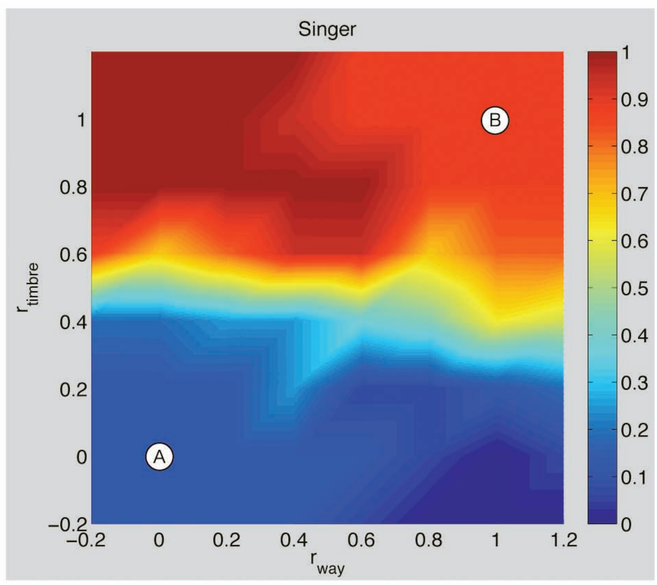
Examples/Applications



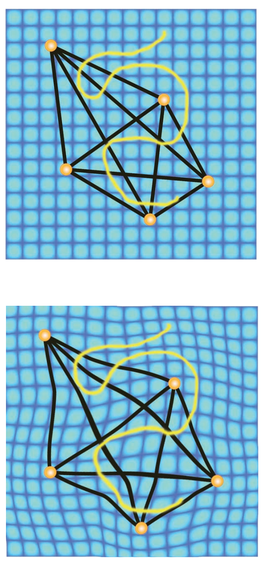
TANDEM-STRAIGHT provides a unified method to extract the excitation source information with supra- and sub- harmonic structures. The baseline performance of the proposed method without any post processing is comparable to available state of the art methods.

Morphing based on STRAIGHT provides means to investigate physical correlates of perceptual attributes in singing. Preliminary tests indicated perceptual dominance of voice timbre in judgement of singers' identity. This finding and a concept of speech texture mapping are basis of a voice conversion solely based on vowel information.





Partial morphing for singing style, timbre and identity



Vowel-based voice conversion