ICSP 2014 Paper Review

Zhehuai Chen

2014.11.13

- Openning and Pictures
- Paper Review (3 fields)
- Some Other Interesting Points

Openning and Pictures

Paper Review

Shanghai Jiao Tong University

- VocaWatcher: Natural Singing Motion Generator for a Humanoid Robot (Keynote)
  - Masataka Goto. AIST

上海交通大学
Shanghai Jiao Tong University
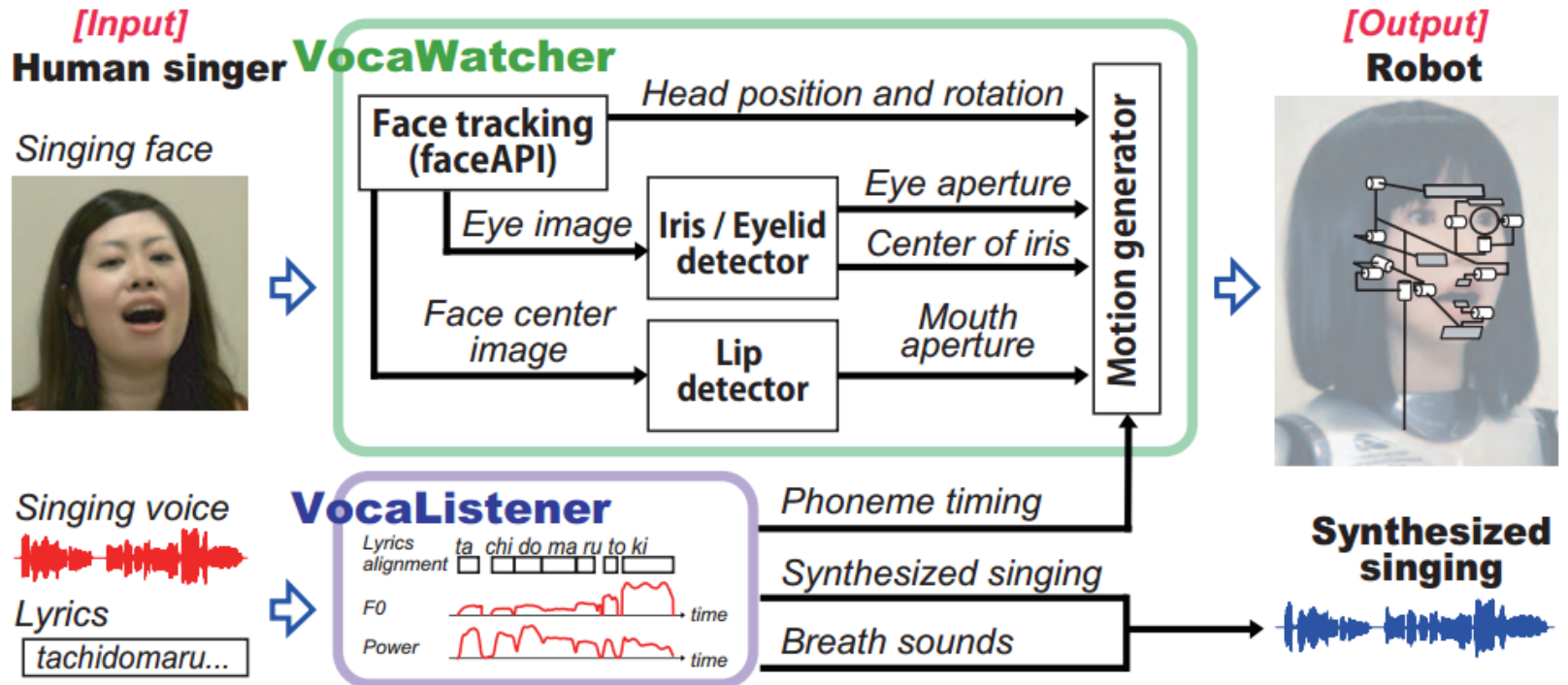
- VocaWatcher: Natural Singing Motion Generator for a Humanoid Robot (Keynote)
  - Masataka Goto. AIST
  - Vocalistener: Imitating Pitch and Dynamics
    - Singing-to-singing: analyzes acoustic features of the input user's singing and directly converts them into synthesis parameters.
    - Phonetic Alignment: use phoneme-level hidden Markov model to adjusted iteratively so that each voiced section of the synthesized singing can be the same as the original voiced section of the target singing.

- VocaWatcher: Natural Singing Motion Generator for a Humanoid Robot (Keynote)
 - Masataka Goto. AIST
  - Vocalistener: Imitating Pitch and Dynamics
  - Vocawatcher: Robot Motion Generation and Imitating Facial Expression
    - Track human's head, eyes, etc.
    - utilize audiobased timing information — Alignment result before

Speech Enhancement

- Noise Identification For Model-Based Speech Enhancement
 - Jiang Wenbin, Ying Rendong, Liu Peilin. SJTU
  - Why to do:
    - the model-based speech enhancement method assumed that "we have prior knowledge of the type of the noise environment such that the correct noise model is used in the enhancement process"
  - Noise Types (8)
    - Babble, exhibition and train noise can be fully solved
  - Noise Signal Feature(4): different shapes in noise
    - MFCC out of 4
  - Noise Model (3): define Spectral Distortion and train
    - GMM out of 3

Speech Enhancement

- A Modified Speech Enhancement Algorithm Using A Universal Speaker Model

  - Li Guo, Wenbin Jiang, and Peilin Liu. SJTU

  - A universal GMM of speakers'MFCCs to caculate weighting factor for noise-reduction(IMCRA & DD method) in frequency-domain

Shanghai Jiao Tong University
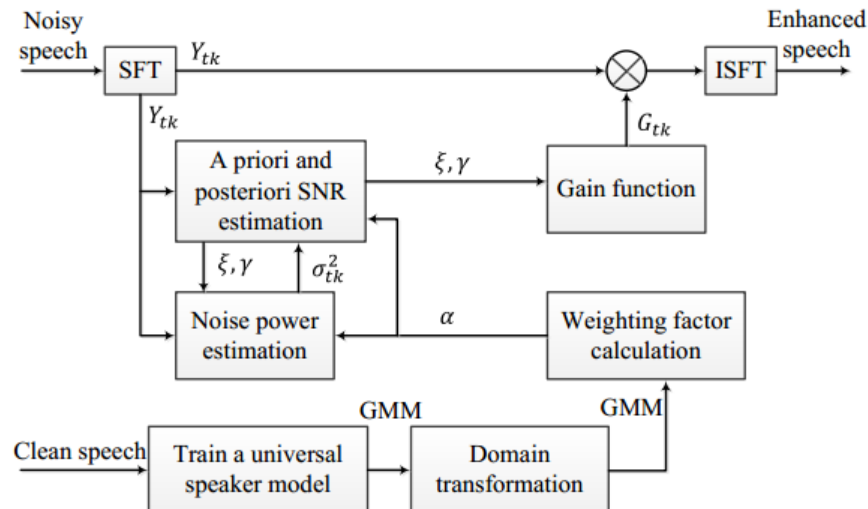
- A Novel Speaker Verification Approach For Certain Noisy Environment
  - Cao Yiming, Jiang Wenbin, Liu Peilin. SJTU
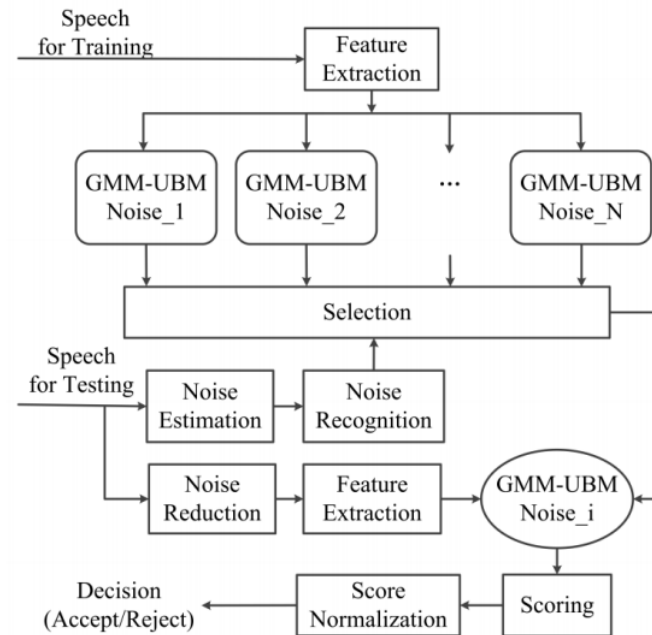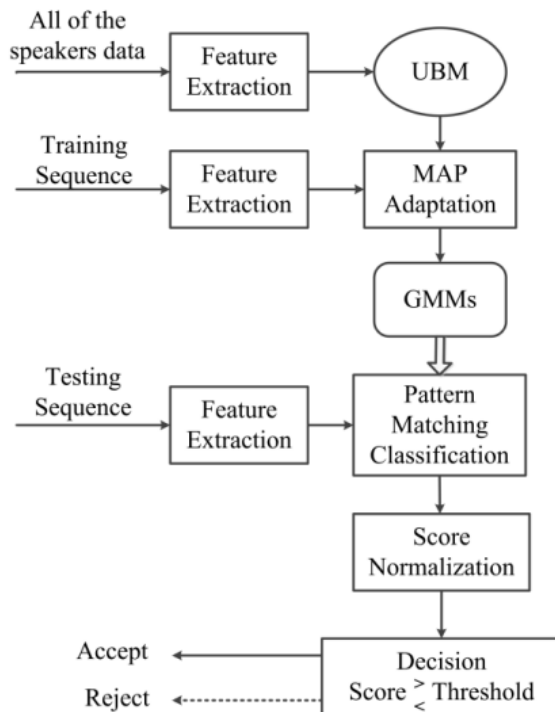
Emotional Speech Synthesis

Multi-Level Prosody And Spectrum Conversion
For Emotional Speech Synthesis

- Zexun Wang, Yibiao Yu. Soochow Uni
- Emotion is labeled by 100 sentences
  F0, energy &duration are considered.

Fig.3.Statistical analysis of F0,energy,duration in four emotions.

- The F0 conversion is in accordance with the order of the syllable, prosodic word and sentence by GMM.
- The energy is modeled by Gamma distribution and transformed in syllable, prosodic word and sentence.
- The **duration** is the same method to energy but by GMM.
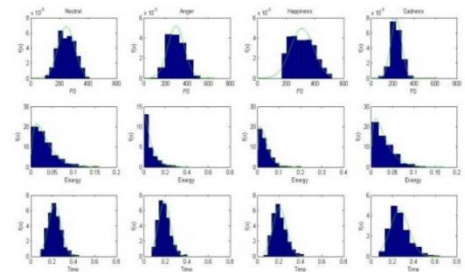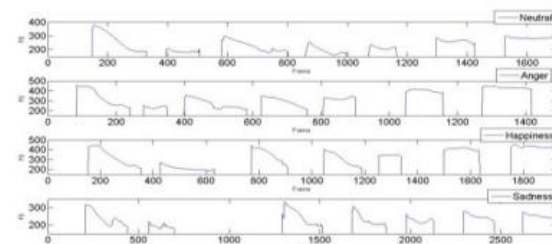- Spectrum is transformed in one leve

Some Other Interesting Points

Some Other Interesting Points

- An Open-Source Gpu-Accelerated Feature Extraction Tool

  - Speed-up between 3 to 18 times

- Comparison Of Performance Between Normal And Whispered Speech In Chinese Isolated Word Recognition

  - Whisper has no pitch frequency
  - the sound level, the tracheal pressure and vital capacity is lower
  - the duration is longer

- Automatic Detection Of Contrastive Word Pairs Using Textual And Acoustic Features
  - Why to:
    - Useful in Speech Synthesis, especially enhance the inteligence of synthetic speech
  - How to:
    - combined use of acoustic features (energy, duration, F0, etc.), part-of-speech (POS) and semantic dissimilarity measure
- Building A Chinese Natural Emotional Audio-Visual Database

Thanks