Abstract

This paper introduce the work done to build an Arabic unit selection voice that could carry emotional information. Three emotional states were covered; normal, sad and questions. An emotional speech classifier was used to enhance the intelligibility of the used recorded speech database. The classification information was employed in the proposed target cost to produce more natural and emotive synthetic speech. The system is evaluated according to the naturalness and emotiveness of the produced speech. The system evaluations show significant increase in the naturalness and emotiveness scores.

References

- Roberto Barra-Chicote, Junichi Yamagishi, Simon King, Juan Manuel Montero, and
  Javier Macias Guarasa. 2010. Analysis of statistical parametric and unit selection speech
  synthesis systems applied to emotional speech, Speech Communication 52(5):394-404
- Brigitte Krenn, Hannes Pirker, Martine Grice, Paul Piwek, Kees van Deemter, Marc
  Schröder, Martin Klesen, and Erich Gstrein. Generation of multimodal dialogue for net
  univie. ac. at/NECA.
- Irene Albrecht, J"org Haber, Kolja Khler, Marc Schröder, and H. -P. Seidel. "May I
  synthesis system. In The Third ESCA Workshop in Speech Synthesis, pages 147-151, Jenolan
  Caves, Australia
- Alan W. Black. 2002. Perfect synthesis for all of the people all of the time. IEEE TTS
- Robert A. J. Clark, Korin Richmond, Simon King. 2007. Multisyn: Open-domain unit
- P. Taylor, A. Black, and R. Caley. 1998. The architecture of the festival speech
  synthesis system. In 3rd ESCA Workshop on Speech Synthesis, pages 147--141, Jenolan
  Caves, Australia.
- Wael Hamza and Mohsen Rashwan. 2000. ‘Concatenative Arabic speech
  182-185, Beijing, China.
- Yong Zhao, Peng Liu, Yusheng Li, Yining Chen and Min Chu. 2006. Acoustics, Speech
  and Signal Processing. ICASSP 2006 Proceedings. 2006 IEEE International Conference on
  (Volume:1 )
  Arabic Diphone Speech Synthesizer in Festival"&apos; . Msc Thesis.
- Al-Haj, H. , Hsiao, R. , Lane, I. , Black, A. , and Waibel, A. "Pronunciation Modeling
  for Dialectal Arabic Speech Recognition"&apos; . ASRU 2009, Merano, Italy.
  Modeling for Dialectal Arabic Speech Recognition. ASRU, page 525-528.
- Anumanchipalli, G. , Prahallad, K. , Black, A. 2011. Festvox: Tools for Creation and
  Analysis of Large Speech Corpora. in Proceedings of Very Large Scale Phonetics Research,
  UPenn, 2011. "
- Stefan Steidl, Tim Polzehl, H. Timothy Bunnell, Ying Dou, Prasanna Kumar
  Muthukumar, Daniel Perry, Kishore Prahallad, Callie Vaughn, Alan W. Black, and Florian
  Metze, Emotion Identification for Evaluation of Synthesized Emotional Speech Prosody
  2012, Shanghai, China.
Keywords
Speech synthesis  Emotion  Festival  Emovoice.

Index Terms
Computer Science  Signal Processing