



LREC 2008 marrakech

COCOSDA/WRITE Workshop Resources for Speech Synthesis

Nick Campbell

NiCT/ATR-SLC
Keihanna Science City,
Kyoto, Japan.
nick@nict.go.jp

Speech Synthesis @ LREC-2008

Speech synthesis, Text-to-speech systems (7 papers)

- Acquiring Pronunciation Data for a Placenames Lexicon in a Less-Resourced Language
- An Automatic Close Copy Speech Synthesis Tool for Large-Scale Speech Corpus Evaluation
- Automatic Phoneme Segmentation with Relaxed Textual Constraints
- Building of a Speech Corpus Optimised for Unit Selection TTS Synthesis
- Corpus and Voices for Catalan Speech Synthesis
- Evaluation of Modules and Tools for Speech Synthesis: the ECESS Framework
- Methodologies for Designing and Recording Speech Databases for Corpus Based Synthesis

SynSIG - The Blizzard Challenge

The SynSIG is very active in this area ...

SynSIG - The Blizzard Challenge (cont.)



[Webmaster login](#)

SynSIG Speech Synthesis Special Interest Group

- [Main Page](#)
- [SynSIG committee](#)
- [Publications](#)
- [Speech Synthesis Workshops](#)
- [Software](#)
- [Education](#)
- [Samples](#)
- [Evaluation](#)
- [Blizzard Challenge](#)
- [Pointers](#)
- [Future events](#)
- [Index](#)
- [Recent changes](#)

Blizzard Challenge 2008

For the Blizzard Challenge 2008, we plan to release two databases: UK English and Mandarin. The results will be presented at a satellite event planned for Interspeech 2008.

Contents [\[hide\]](#)

- 1 [Read these first](#)
- 2 [Data download and license agreements](#)
- 3 [Mailing list](#)
- 4 [Timeline](#)
- 5 [Any questions?](#)
- 6 [Previous challenges](#)

Read these first [\[edit\]](#)

- Download the [Announcement of the Blizzard Challenge 2008](#) in PDF format.
- Read and agree to the [Blizzard Challenge 2008 Rules](#) before participating.

Data download and license agreements [\[edit\]](#)

SynSIG Topics in speech synthesis (for use in calls for papers) (from "http://www.synsig.org/index.php/Future_events")

- * Text processing for speech synthesis
- * Prosody Generation for speech synthesis
- * Speech modeling for speech synthesis applications
- * Signal processing for speech synthesis
- * Concatenative speech synthesis (diphones, polypohones, uni
- * Articulatory synthesis
- * Statistical parametric speech synthesis
- * Voice transformation/conversion/adaptation for speech synt
- * Expressive speech synthesis
- * Multilingual and/or Multimodal Speech Synthesis
- * Text-to-Speech and Content-to-Speech
- * Singing speech synthesis
- * Systems and Applications involving speech synthesis
- * Techniques for assessing synthetic speech quality

What can we conclude from this?

Speech Synthesis now appears to be a mature technology, no longer in need of COCODA support and perhaps not wishing for COCODA assistance in evaluation campaigns.

We have done our job.

The Blizzard Challenge provides direction, data, and evaluations for current speech synthesis research and is widely supported in the community.

The main current research areas include expressivity, small systems, hybrid (stochastic/concatenative) systems, task localisation, and language diversification.

Stuck in the 20th Century? : What speech synthesis can't do yet

speech synthesis *can* already speak well, with a variety of voices, and in a variety of languages
but it can't *talk* yet!

human speakers adjust their voices and speaking styles according to a wide range of factors not just the mood of the speaker, but also the *intentions* of the discourse, and the *strategies* of the dialogue

in particular, we adjust our presentation style and tone-of-voice according to the relationship with the listener, and to the feedback signals that they display throughout the dialogue ...

Emotion @ LREC-2008

Emotions (21 papers)

- Building Affective Lexicons from Specific Corpora for Automatic Sentiment Analysis
- A Multi-Lingual Dictionary of Dirty Words
- Uncertainty Corpus: Resource to Study User Affect in Complex Spoken Dialogue Systems
- On the Role of the NIMITEK Corpus in Developing an Emotion Adaptive Spoken Dialogue System
- Semi-automatic Building Method for a Multidimensional Affect Dictionary for a New Language
- Sentiment Analysis and the Use of Extrinsic Datasets in Evaluation
- Coding Emotional Events in Audiovisual Corpora
- Audio Database in Support of Potential Threat and Crisis Situation Management
- Emotion Recognition from Speech: Stress Experiment

Emotion @ LREC-2008 (cont.)

- Opinion Annotation in On-line Chinese Product Reviews
- Subjective Evaluation of an Emotional Speech Database for Basque
- Annotating Expressions of Opinion and Emotion in the Italian Content Annotation Bank
- Automatic Emotional Degree Labeling for Speakers Anger Utterance during Natural Japanese Dialog
- Estimating Word Phonosemantics
- A Real-World Emotional Speech Corpus for Modern Greek
- Annotating Subjective Content in Meetings
- Resources for Persuasion
- Valentino: A Tool for Valence Shifting of Natural Language Texts
- What is poorly Said is a Little Funny
- A Comparative Cross-Domain Study of the Occurrence of Laughter in Meeting and Seminar Corpora
- Multimodal Spontaneous Expressive Speech Corpus for Hungarian

Where there is perhaps room for improvement

There is considerable interest in “emotional” speech synthesis, and databases of “emotional speech” are being produced, often using actors to perform a small set of semantically unmarked sentences in a wide range of emotions

However, this way of looking at emotions examines *what is happening within the speaker* rather than *what characterises different types of engagement between the speaker & listener* in a discourse.

As a result, except for limited uses such as animated cartoons and fairy-tales, it is likely that the characteristics of speech generated by systems trained on such data will not be appropriate to the wide range of applications in which a speech synthesiser is required to substitute for a human voice.

Some possible areas for advanced research

- localisation - adapting to the various situations
- expressivity - showing engagement and processing an *intentional* level of speech
- collecting voices in a wide range of social situations
- extending hybrid systems (no comment!)
- sharing resources and pooling knowledge

Next Steps in Speech Synthesis

- producing fragmented utterances
- monitoring feedback & adapting to the listener
- listener sensing (attention/comprehension)
- providing feedback utterances and nonverbal speech
- speaker & speaking-style adaptation

Conclusion

- FP7 NoE Proposal - ENLIST - good luck!!
- latest Blizzard - watch this space!
- basic science : we need a deeper understanding of multimodal discourse features and of conversational engagement
- language/dialect, task, & speaker temporal characteristics

See you next year!

thanks

Thank you for your attention