

Concept to Speech Generation Systems

**Proceedings of a Workshop
in conjunction with 35th Annual Meeting of the
Association for Computational Linguistics**

Edited by

Kai Alter, Hannes Pirker, and Wolfgang Finkler

11 July 1997

**Universidad Nacional de Educación a Distancia
Madrid, Spain**

TABLE OF CONTENTS

Organizing and Program Committee	ii
Program Timetable	iii
Introduction to the Workshop	iv
<i>Probabilistic Model of Acoustic / Prosody / Concept Relationships for Speech Synthesis</i>	
Nanette M. Veilleux	1
<i>Message-to-Speech: High Quality Speech Generation for Messaging and Dialogue Systems</i>	
Peter Spyns, Filip Deprez, Luc Van Tichelen, and Bert Van Coile	11
<i>A Compact Representation of Prosodically Relevant Knowledge in a Speech Dialogue System</i>	
Peter Poller and Paul Heisterkamp	17
<i>Integrating Language Generation with Speech Synthesis in a Concept to Speech System</i>	
Shimei Pan and Kathleen McKeown	23
<i>Can Pitch Accent Type Convey Information Status in Yes-No Questions?</i>	
Martine Grice and Michelina Savino	29
<i>Computing Prosodic Properties in a Data-to-Speech System</i>	
M. Theune, E. Klabbers, J. Odijk, and J.R. de Pijper	39
<i>Semantic and Discourse Information For Text-to-Speech Intonation</i>	
Laurie Hiyakumoto, Scott Prevost, and Justine Cassell	47
<i>Looking for the Presence of Linguistic Concepts in the Prosody of Spoken Utterances</i>	
Gerit P. Sonntag and Thomas Portele	57

Program Committee

Robert Bannert
Univ. of Umea, Sweden

John Bateman
GMD Darmstadt, Germany

Mary Beckman
Ohio State Univ., USA

Carlos Gussenhoven
Univ. of Nijmegen, The Netherlands

Bjorn Granstroem
KTH Stockholm, Sweden

Elisabeth Maier
DFKI Saarbrücken, Germany

Scott Prevost
MIT Boston, USA

Mark Steedman
Univ. of Pennsylvania, USA

Organizing Committee

Kai Alter
Austrian Research Institute for AI (OFAI) &
Max-Planck-Institute of Cognitive Neuroscience
alter@cns.mpg.de

Wolfgang Finkler
German Research Center for AI (DFKI)
finkler@dfki.uni-sb.de

Hannes Pirker
Austrian Research Institute for AI (OFAI)
hannes@ai.univie.ac.at

PROGRAM TIMETABLE

9.00 am	Introduction	
9.10 am	Veilleux, N.	<i>Probabilistic Model of Acoustic / Prosody / Concept Relationships for Speech Synthesis</i>
9.45 am	Spyns, P., Deprez, F., Van Tichelen, L., & Van Coile, B.	<i>Message-to-Speech: High Quality Speech Generation for Messaging and Dialogue Systems</i>
10.20 am	Poller, P. & Heisterkamp, P.	<i>A Compact Representation of Prosodically Relevant Knowledge in a Speech Dialogue System</i>
10.55 am	Coffee Break	
11.15 am	Pan, S. & McKeown, K.	<i>Integrating Language Generation with Speech Synthesis in a Concept to Speech System</i>
11.50 am	Grice, M. & Savino, M.	<i>Can Pitch Accent Type Convey Information Status in Yes-No Questions?</i>
12.25 pm	Theune, M., Klabbers, E., Odijk, J., & de Pijper, J.R.	<i>Computing Prosodic Properties in a Data-to-Speech System</i>
1.00 pm	Lunch Break	
3.00 pm	Hiyakumoto, L., Prevost, S., & Cassell, J.	<i>Semantic and Discourse Information for Text-to-Speech Intonation</i>
3.35 pm	Sonntag G.P., & Portele Th.	<i>Looking for the Presence of Linguistic Concepts in the Prosody of Spoken Utterances</i>
4.10 pm	Final Discussion	

Introduction to the Workshop

Traditionally, research on spoken language generation was mainly undertaken within the separate fields of natural language generation and speech synthesis. On the one hand, current generation systems allow for the production of flexible utterances. They may be utilized to overcome the limitations of Human-Computer Interfaces with stereotyped language output. However, they typically neglect aspects of intonation and hand over the resulting text in graphemic form to a speech synthesis component.

On the other hand, current speech synthesizers implement the so-called Text-to-Speech approach. They are able to read aloud unrestricted text. One of the main problems posed by that paradigm is the production of adequate prosody since the written form of a text that is to be articulated is a poor knowledge source. Prosodic features of an utterance are highly dependent on the informational structure, on the linguistic structure, and on the situational context of the utterance.

A tight interaction between generation and synthesis should contribute to enhance the quality of a system's output. A component for speech synthesis may be provided with relevant parameters to compute adequate intonation. A generation system may utilize options of speech synthesis during its decisions of tactical or strategic generation, e.g., reflect the information structure either by intonational cues or via morpho-syntactic variations (e.g. changing of word order).

Concept-to-Speech (CTS) generation, i.e., the production of synthetic speech on the basis of pragmatic, semantic, and discourse knowledge offers a challenging and relatively new field of research in intelligent user interfaces. The questions raised in such an environment range from pragmatics, semantics, and (morpho-)syntax to phonology and phonetics. The modelling of prosody (at symbolic and acoustic level) serves as one of the open questions within this paradigm. Obviously, the development of a CTS system is very demanding. Successful work within the framework of CTS relies on the ability to integrate efforts from a number of disciplines, such as Computational Linguistics, Artificial Intelligence, Cognitive Science, and Signal Processing.

The workshop will provide a forum to bring together researchers from the fields of natural language generation and speech synthesis. The aim of the workshop is to stimulate interchange of innovative ideas and results of diverse aspects of CTS generation in order to bridge the gap between these fields.

Among the challenging aspects of a CTS system, we proposed to address issues of the following list in the first place:

- How can systems for natural language generation be adapted in order to utilize new realization options to the generation process that are offered in the CTS framework?
- How can issues in the time-course of the interleaved process for generation and synthesis (when-to-say) be dealt with? Which requirements on speech synthesis are to be fulfilled in an incremental approach to spoken language production?
- Due to its inherent integrational property, being influenced by a whole number of representational levels, modelling of prosody will be one of the major topics of the workshop.
- How can approaches in the Text-to-Speech tradition to synthesis show their adaptability to Concept-to-Speech?

We invited contributions that provide solutions to any of the topics indicated above or that present innovative applications addressing the abovementioned issues.