

# **Speech and Signals - Aspects of Speech Synthesis and Automatic Speech Recognition**

Dedicated to  
Wolfgang Hess on his 60th birthday

edited by  
Walter F. Sendlmeier

**Frankfurt am Main 2000**

# Contents

<b>Editor's Introduction</b>	<b>13</b>
<b>Louis C.W. Pols: <i>Even better synthesis</i></b>	<b>15</b>
Abstract	15
Personal Introduction	15
Some History	15
From Munich to Bonn	18
ESCA/ISCA	19
Speech Synthesis	19
Evaluation	21
Conclusion	22
References	22
<b>Felix Burkhardt und Walter F. Sendmeier: <i>Verification of Acoustical Correlates of Emotional Speech using Formant-Synthesis</i></b>	<b>27</b>
Abstract	27
Introduction	27
emoSyn: A Speech Synthesizer optimized for Emotional Expression	28
Overview	28
Description of Modifiable Parameters	29
Experiment 1: Systematic Variation of Acoustical Features	30
Generating the Stimuli	30
The Perception Experiment	30
Statistical Analysis and Interpretation of the Results	31
Summary	34
Experiment 2: Further Optimization of Specific Features	34
The Prototypes and their Variations	34
The Perception Experiment	35
Analysis and Discussion of the Results	35
Conclusions	38
Acknowledgements	38
References	38

<b>Eric Keller, Brigitte Zellner-Keller &amp; John Local: <i>A Serial Prediction Component for Speech Timing</i></b>	<b>41</b>
Abstract . . . . .	41
Introduction . . . . .	41
Description of the Serial Dimension . . . . .	42
Serial Dependency . . . . .	43
Method . . . . .	44
Data Bases and Speakers . . . . .	44
Analysis . . . . .	44
Results . . . . .	45
Conclusions . . . . .	46
Acknowledgements . . . . .	47
References . . . . .	48
<b>Gerit P. Sonntag: <i>Evaluation of Speech Synthesis Systems</i></b>	<b>51</b>
Abstract . . . . .	51
Introduction . . . . .	51
The quality of speech . . . . .	53
How to define 'speech quality' . . . . .	53
Different dimensions of speech quality . . . . .	54
Prosody . . . . .	57
PURR — a method for prosody evaluation . . . . .	57
Some experiments carried out with the PURR-method . . . . .	58
Comparison of synthetic voices . . . . .	58
Comparison of human voices . . . . .	59
Analysing prosodic functions . . . . .	59
Conclusion . . . . .	61
References . . . . .	62
<b>Klaus Fellbaum and Bettina Ketzmerick: <i>On the Suitability of Voices for Speech Synthesis</i></b>	<b>65</b>
Introduction . . . . .	65
Survey of related work . . . . .	67
Suitability of voices — auditory part . . . . .	68
Characterisation of voices . . . . .	68
Listening tests . . . . .	69
Acoustical part . . . . .	70
Pitch (fundamental frequency) and its range of variation . . . . .	71
Formant frequencies . . . . .	71
Articulation rate . . . . .	72
Time structures, durations of syllables and sounds . . . . .	72
Energy distributions . . . . .	73
Selected results of the apparative measurements . . . . .	73

Correlation between auditive and acoustical features . . . . .	74
Concluding remarks . . . . .	75
References . . . . .	77
<b>Bernd Möbius: <i>Corpus-Based Speech Synthesis: Methods and Challenges</i></b>	<b>79</b>
Abstract . . . . .	79
Introduction . . . . .	79
Selection algorithm . . . . .	81
Distance measures and weights . . . . .	81
Context clustering . . . . .	83
Phonetic and phonological trees . . . . .	85
Speech database . . . . .	87
Word and syllable concatenation . . . . .	90
Looking ahead . . . . .	91
References . . . . .	91
<b>Barbara Heuft, Thomas Portele, Petra Wagner, Christina Widera, Maria Wolters:</b>	
<i>Perceptual Prominence</i>	<b>97</b>
Abstract . . . . .	97
Motivation . . . . .	97
Labeling and Perception . . . . .	98
Acoustic Correlates . . . . .	101
Prosodic Parameters . . . . .	101
Segmental Parameters . . . . .	104
Functions . . . . .	107
Linguistic Annotation of the BPD . . . . .	107
Part-of-Speech and Syntax . . . . .	108
Focus . . . . .	109
Prediction . . . . .	112
Conclusions . . . . .	113
References . . . . .	114
<b>Rüdiger Hoffman: <i>The Dresden Speech Synthesis System DreSS</i></b>	<b>117</b>
Abstract . . . . .	117
History . . . . .	117
Components of DreSS . . . . .	118
Text processing . . . . .	118
Grapheme-phoneme conversion . . . . .	119
Prosodic control . . . . .	119
Unit selection . . . . .	121
Acoustic synthesis . . . . .	122
Multilinguality of DreSS . . . . .	122
Platforms and hardware support . . . . .	122

DreSS and the internet . . . . .	123
An internet tutorial about DreSS . . . . .	124
A framework for designing speech applications in the web . . . . .	124
Acknowledgement . . . . .	125
References . . . . .	125
<b>Erwin Paulus: <i>Some guidelines for the evaluation of approaches to automatic speech recognition</i></b>	<b>129</b>
Abstract . . . . .	129
Introduction . . . . .	129
Design and performance issues . . . . .	131
Models and training . . . . .	131
Word error rate and word accuracy . . . . .	132
Speech recognition output . . . . .	133
Performance and limitations . . . . .	135
Operating characteristics . . . . .	136
Error rate and training set . . . . .	136
Error rate and complexity . . . . .	137
Error rate and word lattice density . . . . .	138
References . . . . .	139
<b>Ailbhe Ní Chasaide and Christer Gobl: <i>Voice Quality Variation: Analysis and Perception of Affect</i></b>	<b>141</b>
Introduction . . . . .	141
Analysing Voice Quality . . . . .	142
Analytic method . . . . .	142
Source parameters . . . . .	143
Resynthesis . . . . .	145
Affective Correlates of Voice Quality: A Perception Experiment . . . . .	145
Stimuli . . . . .	147
Perception test . . . . .	149
Results of Perception Experiment . . . . .	149
Conclusions . . . . .	152
Acknowledgement . . . . .	154
References . . . . .	154
<b>Miriam Kienast, Walter F. Sendlmeier: <i>Acoustical Analysis of Spectral and Temporal Changes in Emotional Speech</i></b>	<b>157</b>
Abstract . . . . .	157
Introduction . . . . .	157
Material . . . . .	158
Methods and Parameters . . . . .	159
Assimilation and Segment Deletions . . . . .	159

Vowel Formant Analysis . . . . .	159
Spectral Balance . . . . .	160
Results . . . . .	160
Assimilations and Segment Deletions . . . . .	160
Vowel Formant Analysis . . . . .	162
Spectral Balance . . . . .	163
Summary and Discussion . . . . .	164
Acknowledgement . . . . .	167
References . . . . .	167

**Thilo Pfau, Manfred Beham, Wolfgang Reichl, Günther Ruske: *Creating Large Subword Units for Speech Recognition* . . . . . 169**

Abstract . . . . .	169
Preamble . . . . .	169
Introduction . . . . .	170
Choosing appropriate base units . . . . .	170
Construction of a MDS Inventory . . . . .	171
Setup of MDS HMM's . . . . .	172
Experimental Results . . . . .	174
Conclusion . . . . .	175
Acknowledgements . . . . .	176
References . . . . .	176

**A Prof. Dr. Wolfgang Hess – Publications as of December 1998 . . . . . 177**

Books . . . . .	177
Papers in Journals and Conference Proceedings . . . . .	177
Presentations with Written or Electronic Documentation . . . . .	184

**B Authors' list . . . . . 185**