

PROCEEDINGS OF THE 6TH INTERNATIONAL CONFERENCE ON SPEECH PROSODY

Shanghai, May 22-25, 2012

Qiuwu Ma, Hongwei Ding and Daniel Hirst (eds.)

Volume I



同济大学出版社
TONGJI UNIVERSITY PRESS



PROCEEDINGS OF THE 6TH INTERNATIONAL CONFERENCE ON SPEECH PROSODY

Shanghai, May 22-25, 2012

Qiuwu Ma, Hongwei Ding and Daniel Hirst (eds.)

常州大学图书馆
藏书章



同济大学出版社
TONGJI UNIVERSITY PRESS

图书在版编目(CIP)数据

第6届国际韵律大会论文集:Proceedings of the 6th International Conference on Speech Prosody 2012:英文/马秋武,丁红卫,(法)赫丹(Hirst,D.)主编.--上海:同济大学出版社,2012.5

ISBN 978-7-5608-4869-3

I. ①第… II. ①马… ②丁… ③赫… III. ①韵律(语言)—国际学术会议—文集—英文 IV. ①H014-53

中国版本图书馆CIP数据核字(2012)第088466号

第6届国际韵律大会论文集

Proceedings of the 6th International Conference on Speech Prosody 2012

马秋武 丁红卫 Daniel Hirst 主编

Qiuwu Ma, Hongwei Ding and Daniel Hirst(eds.)

责任编辑 赵泽毓 责任校对 张德胜 封面设计 陈益平

出版发行 同济大学出版社 www.tongjipress.com.cn

(地址:上海市四平路1239号 邮编:200092 电话:021-65985622)

经 销 全国各地新华书店

印 刷 同济大学印刷厂

开 本 889 mm × 1240 mm 1/16

印 张 48

字 数 1 540 000

版 次 2012年5月第1版 2012年5月第1次印刷

书 号 ISBN 978-7-5608-4869-3

定 价 368.00 元

本书若有印装质量问题,请向本社发行部调换 版权所有 侵权必究

Table of Contents

Tuesday

K1 Analysis-by-Synthesis in Prosody Research

<i>Rüdiger Hoffmann</i>	1
-------------------------------	---

PS1 Speech Processing and Tools; Descriptive and Cross-linguistic Studies

PS1A Speech Processing and Tools

PS1A-1 Representing the Prosodic Context of Words using Gaussian Mixture Models <i>Shreyas A. Karkhedkar, Nigel G. Ward</i>	7
PS1A-2 Prosody Modification for Vocoder Based on Amplitude Spectrum of Residual Signal <i>Zhengqi Wen, Jianhua Tao</i>	11
PS1A-3 ProZed: A speech prosody analysis-by-synthesis tool for linguists <i>Daniel Hirst</i>	15
PS1A-4 SPEECH Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody <i>Brigitte Bigi, Daniel Hirst</i>	19
PS1A-5 Tone Generation by Maximizing Joint Likelihood of Syllabic HMMs for Mandarin Speech Synthesis <i>Xingyu Na, Chaomin Wang, Xiang Xie, Jingming Kuang, Yaling He</i>	23
PS1A-6 Automatic segmentation of English words using phonotactic and syllable information <i>Raymond W. M. Ng, Keikichi Hirose</i>	27
PS1A-7 Framework For Consistent Speech Databases <i>Sören Wittenberg, Rüdiger Hoffmann</i>	31
PS1A-8 Perception of Japanese Prosodical Phonemes through use of a Bone-conducted Ultrasonic Hearing-aid <i>Takayuki Kagomiya, Seiji Nakagawa</i>	35
PS1A-9 Role of pitch slope and duration in synthesized Mizo tones <i>D. Govind, Priyankoo Sarmah, S. R. Mahadeva Prasanna</i>	39
PS1A-10 Automatic detection of voice creak <i>Philippe Martin</i>	43
PS1A-11 Multi methods pitch tracking <i>Philippe Martin</i>	47
PS1A-12 Perceptual and acoustic correlates of speech in a bilateral cochlear implant user <i>Aline Neves Pessoa, Beatriz CAC Novaes, Sandra Madureira, Zuleica A Camargo</i>	51
PS1A-13 PROSOTRAN: a tool to annotate prosodically non-standard data <i>Katarina Bartkova, Elisabeth Delais-Roussarie, Fabian Santiago-Vargas</i>	55
PS1A-14 Multi-level Exemplar-Based Duration Generation for Expressive Speech Synthesis <i>Mohamed Abou-Zleikha, Éva Székely, Peter Cahill, Julie Carson-Berndsen</i>	59
PS1A-15 Automatisatation of intonation modelling and its linguistic anchoring <i>Uwe D. Reichel</i>	63

PS1A-16 Using Noisy Speech to Study the Robustness of a Continuous F0 Modelling Method in HMM-based Speech Synthesis <i>Kalu U. Ogbureke, João P. Cabral, Julie Carson-Berndsen</i>	67
PS1A-17 Automatic Analysis of The Prosodic Variations in Parkinsonian Read and Semi-Spontaneous Speech <i>Céline De Looze, Alain Ghio, Stefan Scherer, Gilles Pouchoulin, François Viallet</i>	71
PS1A-18 ProsoDyn: a graphical representation of macroprosody for phonostylistic ambiance change detection <i>Jean-Philippe Goldman</i>	75
PS1B Descriptive and Cross-linguistic Studies	
PS1B-1 Pitch accent, eyebrows and context — effects on the exhaustivity of answers? <i>Charlotte Wollermann, Bernhard Schröder, Ulrich Schade</i>	79
PS1B-2 Alignment of Intonational Events in German and Brazilian Portuguese — a Quantitative Study <i>Hansjörg Mixdorff, Plínio A. Barbosa</i>	83
PS1B-3 The Effect of Language Attrition and Tone Sandhi on Taiwanese Tonal Processing <i>Chia-Hsin Yeh, Jung-Yueh Tu</i>	87
PS1B-4 The Effects of Prosodic Context and Word Position on Gupapuyngu Vowels <i>Simone Graetzer, John Hajek, Janet Fletcher</i>	91
PS1B-5 The Falsetto Tones of the Dialects in Hubei Province <i>C. Y. Wang, C. J. Tang</i>	95
PS1B-6 Cross-Linguistic Perception of Intonation by Mandarin and Cantonese Listeners <i>Bo Robert Xu, Peggy Mok</i>	99
PS1B-7 Prosodic Word in Chengdu Dialect <i>Zuxuan Qin</i>	103
PS1B-8 Disambiguation of Tonally Unspecified Mandarin Syllables <i>Christopher Huff, Elinor Payne</i>	107
PS1B-9 A Global Perspective on Chinese Intonation <i>Wu Lihan, Ma Qiuwu</i>	111
PS1B-10 Voice quality and gender: some insights on correlations between perceptual and acoustic dimensions <i>Zuleica Camargo, Sandra Madureira, Aline Neves Pessoa, Luiz Carlos Rusilo</i>	115
PS1B-11 Towards the Mo Piu Tonal System: First Results on an Undocumented South-Asian Language <i>Geneviève Caelen-Haumont</i>	119
PS1B-12 Functions of word order and intonation in information structuring in Puyuma <i>Anastasia Karlsson, Arthur Holmer</i>	123
PS1B-13 Sport in the media: a contrasted study of three sport live media reports with semi-automatic tools <i>Stéphanie Audrit, Tea Pršir, Antoine Auchlin, Jean-Philippe Goldman</i>	127
PS1B-14 The Autosegmental-Metrical Prosodic Structure: not fit for French? <i>Philippe Martin</i>	131
PS1B-15 Introducing Nasalization in Nkonya <i>Rogers Krobea Asante</i>	135

OS1 Prosody and Speech Processing

OS1-1 Prosody-dependent Acoustic Modeling for Mandarin Speech Recognition <i>Tzu-Hsuan Chiu, Chen-Yu Chiang, Yuan-Fu Liao, Jyh-Her Yang, Yih-Ru Wang, Sin-Horng Chen</i>	139
--	-----

OS1-2 Exploiting Time and Frequency Domain Measures for Precise Voice Source Parameterisation	
<i>John Kane, Irena Yanushevskaya, Ailbhe Ní Chasaide, Christer Gobl</i>	143
OS1-3 Specific contribution of tonal and duration cues to the syntactic parsing of French	
<i>Amandine Michélas, Mariapaola D'Imperio</i>	147
OS1-4 The Auditory Kappa Effect in a Speech Context	
<i>Alejna Brúgos, Jonathan Barnes</i>	151
OS1-5 Investigating Holistic Measures of Speech Prosody through Ratings of Filtered Speech	
<i>Aliel Cunningham</i>	155
OS1-6 On The Definition of the Word "Segmental"	
<i>Mark Hasegawa-Johnson, Elabbas Benmamoun, Eiman Mustafawi, Mohamed Elmahdy, Rehab Duwairi</i>	159

SS1 Control of Prosodic Features for Expressive Speech Synthesis with Less Speech Corpus

SS1-1 Emotional Voice Conversion for Mandarin using Tone Nucleus Model — Small Corpus and High Efficiency	
<i>Miaomiao Wang, Miaomiao Wen, Keikichi Hirose, Nobuaki Minematsu</i>	163
SS1-2 Assign Stress for Interrogative Sentences via Syntax Structure Mapping	
<i>Ya Li, Xuefei Liu, Xiaoying Xu, Jianhua Tao</i>	167
SS1-3 Fundamental Frequency Contour Reshaping in HMM-based Speech Synthesis and Realization of Prosodic Focus Using Generation Process Model	
<i>Keikichi Hirose, Hiroya Hashimoto, Jun Ikeshima, Nobuaki Minematsu</i>	171
SS1-4 Statistical Approach to Fujisaki-Model Parameter Estimation from Speech Signals and Its Quantitative Evaluation	
<i>Kota Yoshizato, Hirokazu Kameoka, Daisuke Saito, Shigeki Sagayama</i>	175
SS1-5 Making Sense of Variations: Introducing Alternatives in Speech Synthesis	
<i>Nicolas Obin, Christophe Veaux, Pierre Lanchantin</i>	179
SS1-6 Analysis of Voice Stress in Call Centers Conversations	
<i>Grazyna Demenko, Magdalena Jastrzębska</i>	183

Wednesday

K2 Modeling Prosody Variations for Communicative Speech and the Second Language towards Trans-disciplinary Scientific Understanding	
<i>Yoshinori Sagisaka</i>	187

OS2 Prosody and First and Second Language Acquisition

OS2-1 A Longitudinal Study of Prosodic Exaggeration in Child-directed Speech	
<i>Soroush Vosoughi, Deb Roy</i>	194
OS2-2 Effects of Learners' Language Transfer on Native Listeners' Evaluation of the Prosodic Naturalness of Japanese Words	
<i>Shuhei Kato, Greg Short, Nobuaki Minematsu, Keikichi Hirose</i>	198
OS2-3 Children's Acquisition of Mandarin Tone 3 Sandhi in Flat Structures	
<i>Chiung-Yao Wang, Cristina Schmitt, Yen-Hwei Lin</i>	202
OS2-4 The Activation of Segmental and Tonal information in Semantic Retrieval among Chinese-English Bilingual Readers	
<i>Chuchu Li, Candice Y. Lin, Min Wang, Nan Jiang</i>	206

PS2 Language Interaction and Acquisition; Prominence, Focus and Emphasis

PS2A Language Interaction and Acquisition

- PS2A-1 Multi-Stage Feature Normalization for Robust German Stressed/Unstressed Syllable Classification**
Yuan-Fu Liao, Yan-Ting Chen, Jhen-Lun Huang 210
- PS2A-2 Accuracy of Japanese pitch accent rises during and after shadowing training**
Rongna A, Ryoko Hayashi 214
- PS2A-3 A Study of Tone Acquisition in a Mandarin-speaking Child under Three**
Huiqin Ma, Xiaoxiang Chen, Xin Liu 218
- PS2A-4 Local vs. Global Prosodic Cues: Effect of Tones on Attitudinal Prosody in Cross-Perception of Vietnamese by French**
Dang-Khoa Mac, Véronique Aubergé, Eric Castelli, Albert Rilliard 222
- PS2A-5 A Comparison Study on F0 Distribution of Tone 2 and Tone 3 in Mandarin Disyllables by Native Speakers and Japanese Learners**
Ting Zou, Jinsong Zhang, Wen Cao 226
- PS2A-6 Lexical Tone Perception in Early Infancy**
Ao Chen, René Kager 230
- PS2A-7 A Corpus Study of Native and Non-native Vowel Quality**
Chen-huei Wu, Chilin Shih 234
- PS2A-8 A Preliminary Investigation of the Effects of Study Tours on L2 Speech Rhythm**
Donald White, Peggy Mok 238
- PS2A-9 On Non-Native Speaker Prosody: Identifying “Just-Noticeable-Differences” of Speaker-Ethnicity**
Richard Todd 242
- PS2A-10 Prosodic Features of the Chinese EFL Learners as They Express New Information**
Meng Wang, Xia Wang, Yuan Jia, Honghua Zhai 246
- PS2A-11 A Comparative Study on Accentual Structure between Spanish Learners of French Interlanguage and French Native Speakers**
M^a Angeles Barquero Armesto 250
- PS2A-12 Native-like Duration Ratio of Stressed vs. Unstressed Syllables through Visualizing Prosody**
Markus Rude 254

PS2B Prominence, Focus and Emphasis

- PS2B-1 Say it again, Sam! The prosodic profiles of emphatic reduplication in German**
Oliver Niebuhr, Patrycja Jarzabkowska, Uta Lorenz, Christoph Schulz, Farhod Sodigov 258
- PS2B-2 Hocus Focus — How prosodic profiles of contrastive focus emerge and change in different elicitation contexts**
Karin Görs, Oliver Niebuhr 262
- PS2B-3 An Acoustic Study of Penultimate Accentuation in Three Varieties of French**
Sandra Schwab, Mathieu Avanzi, Jean-Philippe Goldman, Pascal Montchaud, Isabelle Racine 266
- PS2B-4 Extra-Sentential Elements, Prosodic Restructuring, and Information Structure. A Study of Clitic-Left Dislocation in Spontaneous French**
Mathieu Avanzi, Lisa Brunetti, Cédric Gendrot 270
- PS2B-5 Revisiting focus prosody in Japanese**
Albert Lee, Yi Xu 274
- PS2B-6 Cross-Language Prominence Detection**
Andrew Rosenberg, Erica Cooper, Rivka Levitan, Julia Hirschberg 278

PS2B-7 Prosodic Effects on Garden-path Sentences	282
<i>Li-Hsin Ning, Chilin Shih</i>	
PS2B-8 Prosodic encoding and perception of focus in Tibetan (Anduo Dialect)	286
<i>Ling Wang, Bei Wang, Yi Xu</i>	
PS2B-9 Accentuation as a Cue for Speaker Adaptation	290
<i>Constantijn Kaland, Emiel Krahmer, Marc Swerts</i>	
PS2B-10 Temporal Characteristics of Emphasis in Continuous Speech	294
<i>Chiu-Yu Tseng, Chao-Yu Su</i>	
PS2B-11 A Prosodic Account of Italian Exclamative Sentences: a Gating Test	298
<i>Patrizia Sorianello</i>	
PS2B-12 Using Map Tasks to Investigate the Effect of Contrastive Focus on the Mandarin Alveolar-Retroflex Contrast	302
<i>Yung-hsiang Shawn Chang, Chilin Shih</i>	
PS2B-13 Compensation for coarticulation in prosodically weak words	306
<i>Felicitas Kleber, Jonathan Harrington, Ulrich Reubold, Jessica Siddins</i>	
PS2B-14 Deaccentuation in Hungarian and its logical background	310
<i>Katalin Mády</i>	
PS2B-15 On the Normalization of Syllable Prominence Ratings	314
<i>Christopher Sappok, Denis Arnold</i>	
PS2B-16 Determining prominence and prosodic boundaries in Korean by non-expert rapid prosody transcription	318
<i>Hie-Jung You</i>	
PS2B-17 Modeling speaker variation in cues to prominence using the Bayesian information criterion	322
<i>Tim Mahrt, Jennifer Cole, Margaret Fleck, Mark Hasegawa-Johnson</i>	
PS2B-18 Phonetic Realization of Prosodic Focus in Persian	326
<i>Mortaza Taheri Ardali, Yi Xu</i>	
PS2B-19 Do native Spanish speakers transfer accentual acoustic properties from Spanish to French L2?	330
<i>Sandra Schwab</i>	

SS2 Prosody and Language in Contact

SS2OSA Prosody in Language Acquisition

SS2OSA-1 The Role of Lexical Knowledge and Stress Cues in Segmentation in Second Language Learners of English	334
<i>Candise Y. Lin, Min Wang</i>	
SS2OSA-2 Acquiring Phrasing and Intonation in French as Second Language: the Case of Yes-No Questions Produced by Mexican Spanish Learners	338
<i>Fabián Santiago Vargas, Elisabeth Delais-Roussarie</i>	
SS2OSA-3 The Effect of Language Attrition on Low Level Tone in Hakka	342
<i>Chia-Hsin Yeh, Chi-Jui Lu</i>	
SS2OSA-4 Non-native Tone Perception from Infant to Adult: How Consistent and Flexible is it?	346
<i>Liquan Liu, René Kager</i>	

SS2PS Prosody in Language Acquisition and Language Contact

SS2PS-1 The role of the acoustic correlates of stress in the perception of Spanish accentual contrasts by French speakers	350
<i>Sandra Schwab, Joaquim Llisterri</i>	

SS2PS-2 Tone Sandhi Patterns of Zhenjiang Dialect	
<i>ChunAn Qiu</i>	354
SS2PS-3 Mandarin-accented Fall, Rise and Fall-Rise F0 contours in Dutch	
<i>Xuliang He, Judith Hanssen, Vincent J. van Heuven, Carlos Gussenhoven</i>	358
SS2PS-4 Evaluation of Benefits from a Computer-Aided Pronunciation Training System for German Learners of Mandarin Chinese	
<i>Hue San Do, Hussein Hussein, Hansjörg Mixdorff, Oliver Jokisch, Hongwei Ding, Qianying Gao, Si Wei, Guoping Hu</i>	362
SS2PS-5 The perception of speech and non-speech tones by tone and non-tone language listeners	
<i>Zhen Qin, Peggy Mok</i>	366
SS2PS-6 A Study on College Non-English Majors' Acquisition of English Stress Placement	
<i>Xiaoxiang Chen, Hongmin Zhang, Xiaojing Xie</i>	370
SS2PS-7 A Phonetic Investigation of Intonational Foreign Accent in Mandarin Chinese Learners of German	
<i>Hongwei Ding, Oliver Jokisch, Rüdiger Hoffmann</i>	374
SS2PS-8 Perception and Production of Mandarin Disyllabic Tones by German Learners	
<i>Hongwei Ding</i>	378
SS2PS-9 YN questions in L2 Italian of Dutch speaking learners	
<i>Claudia Crocco, Jens Baele</i>	382
SS2PS-10 L1 Prosodic transfer and priming effects: A quantitative study on semi-spontaneous dialogues	
<i>Giuseppina Turco, Michele Gubian</i>	386
SS2PS-11 How Do Pauses Reveal Linguistic Plans by L2 Learners of Chinese?	
<i>Bei Yang</i>	390
SS2PS-12 Phonetic and Phonological Realization of Narrow Focus in English Declarative Sentences by Zhenjiang EFL Learners	
<i>Hu Na, Jia Yuan, Liu Bin</i>	394
SS2PS-13 Prosody Perception in L1 and L2	
<i>Irina Nesterenko</i>	398
SS2PS-14 Perception of Focus Prosody in the English speech of Bangou speakers	
<i>Dorothea J. Hackman</i>	402
SS2PS-15 Regional Variations of Speech Rhythm in French: <i>In Search of Lost Times</i>	
<i>Nicolas Obin, Mathieu Avanzi, Guri Bordal, Alice Bardiaux</i>	406
SS2PS-16 Post-Focus Compression in Second Language Mandarin	
<i>Ying Chen, Susan Guion-Anderson, Yi Xu</i>	410
SS2PS-17 Language Contact and Prosodic Interference: Nuclear Configurations in Occitan and French Statements of the Obvious	
<i>Rafèu Sichel-Bazin, Carolin Buthke, Trudel Meisenburg</i>	414
SS2PS-18 Corsican French Questions: is There a Prosodic Transfer from Corsican to French and How to Highlight it?	
<i>Philippe Boula de Mareüil, Albert Rilliard, Paolo Mairano, Jean-Pierre Lai</i>	418
SS2PS-19 English Lexical Stress and Spoken Word Recognition in Korean Learners of English	
<i>Jeonghwa Shin, Shari R. Speer</i>	422
SS2PS-20 Productions of "continuation contours" by French speakers in L1 (French) and L2 (English)	
<i>Katarina Bartkova, Anne Bonneau, Vincent Colotte, Mathilde Dargnat</i>	426
SS2PS-21 Dialect Adaptation and Two Dimensions of Tune	
<i>James Sneed German</i>	430

SS2PS-22 Prosodic Similarities in French Spoken in the Mascareignes <i>Philippe Martin</i>	434
SS2OSB Prosody in Language Contact	
SS2OSB-1 Prosodic Transfer from Italian to Spanish: Rhythmic Properties of L2 Speech and Argentinean <i>Porteño</i> <i>Ariadna Benet, Christoph Gabriel, Elena Kireva, Andrea Pešková</i>	438
SS2OSB-2 Variations in the Realization of the French Accentual Phrase in the Light of Language Contact <i>Guri Bordal, Mathieu Avanzi, Nicolas Obin, Alice Bardiaux</i>	442
SS2OSB-3 Markedness in the prosody of contact varieties of South African English <i>Sabine Zerbian</i>	446

Thursday

OS3 Prosody and Emotion; Tonal Aspects of Prosody

OS3A Prosody and Emotion

OS3A-1 Prosodic and Acoustic Features of Emotional Speech in Taiwan Mandarin <i>Hsin-Yi Lin, Janice Fon</i>	450
OS3A-2 Effects of Emotion on the Lower Lip Movements at Phrase Boundaries <i>Sungbok Lee, Shrikanth Narayanan</i>	454

OS3B Tonal Aspects of Prosody

OS3B-1 Sonority and Syllable Weight Determine Tonal Association in Tashlhiyt Berber <i>Timo B. Röttger, Rachid Ridouane, Martine Grice</i>	458
OS3B-2 Effects of Tone Merging and Musical Training on Cantonese Tone Perception <i>Peggy Mok, Donghui Zuo</i>	462

PS3 Duration and Rhythm; Pitch and Boundaries

PS3A Duration and Rhythm

PS3A-1 Syllabic Intensity Variations as Quantification of Speech Rhythm: Evidence from Both L1 and L2 <i>Lei He</i>	466
PS3A-2 Significance of Glottal Activity Detection for Duration Modification <i>D. Govind, S. R. Mahadeva Prasanna, B. Yegnanarayana</i>	470
PS3A-3 Rhythm Typology of Brazilian Portuguese dialects <i>Alexsandro Rodrigues Meireles, Viviany de Paula Gambarini</i>	474
PS3A-4 Effect of topic structure and Sentence length on pause in Mandarin Chinese: Comparing female with male speakers <i>Qian Wu, Bei Wang, Xiaxia Zhang</i>	478
PS3A-5 Testing the Perception of Speech Rhythm on Natural and Artificial Stimuli <i>Paolo Mairano, Antonio Romano</i>	482
PS3A-6 Speech Rhythm and Timing of a Speech-impaired Speaker <i>Tae-Jin Yoon, Karin Humphreys</i>	486
PS3A-7 A Taste of Prosody: Possible Effects of the Word-Initial Singleton-Geminate Contrast on Post-Consonantal Vowel Duration in Kelantan Malay <i>Hilmi Hamzah, John Hajek, Janet Fletcher</i>	490

PS3A-8 Significance of Duration in the Prosodic Analysis of Assamese	
<i>D. Govind, S. Mahanta, S. R. Mahadeva Prasanna</i>	494
PS3A-9 Temporal Organization of Syllable Production in the Speech of Cantonese-speaking Children	
<i>Wai-Sum Lee, Eric Zee</i>	498
PS3A-10 Rhythm and Tempo in Slovak	
<i>Štefan Beňuš, Juraj Šimko</i>	502
PS3A-11 A Study on the Metrical Structure of Music with Similarity Experiments	
<i>Stephan Hübler, Rüdiger Hoffmann</i>	506
PS3A-12 A preliminary analysis of speech rhythm in three varieties of Irish (Gaelic)	
<i>Amelie Dorn, Maria O'Reilly, Ailbhe Ní Chasaide</i>	510
PS3A-13 The Neutralization of Tone-related Duration Differences in Sung Cantonese	
<i>Murray Schellenberg</i>	514
PS3A-14 Stress and phonemic length in the perception of Slovak vowels	
<i>Štefan Beňuš, Katalin Mády</i>	516
PS3A-15 Rhythm in Mandarin Chinese and Italian: the Role of Sentence Accents	
<i>Pier Marco Bertinetto, Chiara Bertini, Na Zhi</i>	520
PS3A-16 Effects of temporal chunking on speech recall	
<i>Annie C. Gilbert, Victor J. Boucher, Boutheina Jemel</i>	524
PS3A-17 Effects of frequency, repetition and prosodic location on ambiguous Mandarin word production	
<i>Seth Wiener, Shari R. Speer, Claire Shank</i>	528
PS3A-18 The effects of speaking rate and intonation on the duration of tones in Mandarin Chinese	
<i>Jiahong Yuan</i>	532
PS3A-19 Prosodic Typology Revisited: Adding Macro-Rhythm	
<i>Sun-Ah Jun</i>	535
PS3A-20 Variability of speech rhythm in synchronous speech	
<i>Volker Dellwo, Daniel Friedrichs</i>	539
 PS3B Pitch and Boundaries	
PS3B-1 Effects of topic structure and syntax on boundary pitch variations in Standard Chinese	
<i>Xiaohong Yang, Yufang Yang</i>	543
PS3B-2 Conceptual Planning in Conversational Mandarin: Pitch Variation in Prosodic Phrasing	
<i>Alvin C.-H. Chen, Shu-Chuan Tseng</i>	547
PS3B-3 F0 as the Dominant Cue to Accent Location in Persian	
<i>Vahideh Abolhasanzadeh</i>	551
PS3B-4 A Quantitative Study on Information Contribution of Prosody Phrase Boundaries in Chinese Speech	
<i>Jinsong Zhang, Wei Li, Yanlu Xie, Wen Cao</i>	555
PS3B-5 A Forced-alignment-based Study of Declarative Sentence-ending "ta" in Korean	
<i>Tae-Jin Yoon, Yoonjung Kang</i>	559
PS3B-6 Positional Constraints on the Initial Rise in French	
<i>Cristel Portes, Mariapaola D'Imperio, Leonardo Lancia</i>	563
PS3B-7 Priming Effects of Tones and Segments in Lexical Processing in Mandarin	
<i>Lan Shuai, Bin Li, Tao Gong</i>	567
PS3B-8 Prosodic Similarity — Evidence from an Imitation Study	
<i>Hansjörg Mixdorff, Jennifer Cole, Stefanie Shattuck-Hufnagel</i>	571

PS3B-9	The Role of F_0 in Mongolian Stress	
	<i>Yumei Sang, Philippe Martin</i>	575
PS3B-10	Contour Patterns of Chinese Intonational Phrases	
	<i>Chen Hu</i>	579
PS3B-11	Pitch Level Variation in French: Categorical or Gradient?	
	<i>David Le Gac, Hiyon Yoo, Katarina Bartkova</i>	583
PS3B-12	The Prosody of Algebra and the Algebra of Prosody: Prosodic Disambiguation of Read Mathematical Formulae	
	<i>Michael Phelan</i>	587
PS3B-13	Emotional Expressiveness of Successive Addition Boundary Tone in Mandarin Chinese	
	<i>Aijun Li, Qiang Fang, Jianwu Dang</i>	591
PS3B-14	Prosodic Boundaries and Prosodic Word in Chengdu Dialect: a Durational Perspective	
	<i>Zuxuan Qin</i>	595

OS4 Prosody and Regional Dialects

OS4-1	Durational Cues at Discourse Boundaries in Taiwan Southern Min	
	<i>Sheng-Fu Wang, Janice Fon</i>	599
OS4-2	Speech Prosody of French Regional Varieties	
	<i>Mathieu Avanzi, Nicolas Obin, Alice Bardiaux, Guri Bordal</i>	603
OS4-3	Rhythmic Variability in Swiss German Dialects	
	<i>Adrian Leemann, Volker Dellwo, Marie-José Kolly, Stephan Schmid</i>	607
OS4-4	Syllable Boundary Effect: Temporal Entrainment in Overlapped Speech	
	<i>Marcin Włodarczak, Juraj Šimko, Petra Wagner</i>	611

SS3 Prosodic Annotation: Past Present and Future

SS3OS-1	Providing linguists with better tools: Daniel Hirst's contribution to prosodic annotation	
	<i>Céline De Looze, Na Zhi, Cyril Auran, Hyong-Sil Cho, Sophie Herment, Irina Nesterenko</i>	615
SS3OS-3	Form versus Function — Prosodic Annotation and Modeling go Hand in Hand	
	<i>Hansjörg Mixdorff</i>	621

Friday

K4 Neural Specializations for Pitch in Tonal Languages

	<i>Jackson T. Gandour</i>	624
--	---------------------------	-----

OS5 Prosody and Perception

OS5-1	The Role of Amplitude Envelope in Cantonese Lexical Tone Perception: Implications for Cochlear Implants	
	<i>Yining V. Zhou, Brett A. Martin</i>	629
OS5-2	Segment Duration and Vowel Quality in German Lexical Stress Perception	
	<i>Klaus J. Kohler</i>	633

OS5-3 The Effect of Min Proficiency on Production and Perception of Tones in Taiwan Mandarin	
<i>Yi-Hsuan Huang, E-Chin Wu, Janice Fon</i>	637
OS5-4 Asymmetries Between Production, Perception and Comprehension of Focus Types in Japanese	
<i>Hyun Kyung Hwang</i>	641

PS4 Style, Attitude and Emotion; Syntax, Narrative, Speech Act and Discourse

PS4A Style, Attitude and Emotion

PS4A-1 On the Use of Melodic Patterns as Prosodic Correlates of Emotion in Spanish	
<i>Juan-María Garrido, Yesika Laplaza, Montserrat Marquina</i>	645
PS4A-2 Cross-Linguistic Cross-modality Perception of English Sad and Happy Speech	
<i>Caroline Menezes, Donna Erickson, Jonghye Han</i>	649
PS4A-3 Perceptual Study for Emotional Speech of Mandarin Chinese	
<i>Ting Wang, Hongwei Ding, Wentao Gu</i>	653
PS4A-4 Relative Importance of Different Components of Speech Contributing to Perception of Emotion	
<i>P. Gangamohan, V. K. Mittal, B. Yegnanarayana</i>	657
PS4A-5 Emotion Identification for Evaluation of Synthesized Emotional Speech	
<i>Stefan Steidl, Tim Polzehl, H. Timothy Bunnell, Ying Dou, Prasanna Kumar Muthukumar, Daniel Perry, Kishore Prahallad, Callie Vaughn, Alan W. Black, Florian Metze</i>	661
PS4A-6 How Quickly Do Listeners Recognize Emotional Prosody in Their Native versus a Foreign Language?	
<i>Marc D. Pell, Jessica Robin, Silke Paulmann</i>	665
PS4A-7 When Brain Differentiates Happy from Neutral in Prosody?	
<i>Xuhai Chen, Yufang Yang</i>	669
PS4A-8 The Influence of Speaking Style on Lexical f_0 Profiles in French	
<i>Rena Nemoto, Martine Adda-Decker, Jacques Durand</i>	673
PS4A-9 Prosodic analysis of Brazilian Portuguese attitudes	
<i>Albert Rilliard, João Antônio de Moraes, Donna Erickson, Takaaki Shochi</i>	677
PS4A-10 Extraction of paralinguistic information carried by mono-syllabic interjections in Japanese	
<i>Carlos Toshinori Ishi, Hiroaki Hatano, Norihiro Hagita</i>	681
PS4A-11 Do You Hear My Attitude? Prosodic Perception of Social Affects in Mandarin	
<i>Yan Lu, Véronique Aubergé, Albert Rilliard</i>	685
PS4A-12 Psychoacoustic Abilities as Predictors of Vocal Emotion Recognition in Autism	
<i>Eitan Globerson, Noam Amir, Michal Lavidor, Liat Kishon-Rabin, Ofer Golan</i>	689
PS4A-13 Usual voice quality features and glottal features for emotional valence detection	
<i>Marie Tahon, Gilles Degottex, Laurence Devillers</i>	693
PS4A-14 Acoustic and phonetic differences in laughter of male children and adults	
<i>Caroline Menezes</i>	697

PS4B Syntax, Narrative, Speech Act and Discourse

PS4B-1 Declarative and Interrogative Intonations by Brain-damaged Speakers of Uygur and Mandarin Chinese	
<i>Xinlu Yang, Jie Liang</i>	701
PS4B-2 Perception of Spontaneous Narrative Structure	
<i>Miguel Oliveira, Jr., Regina Cruz, Ebson Wilkerson Silva</i>	705

PS4B-3	A phase-based account of Mandarin third tone sandhi domains	
	<i>Ya Li</i>	709
PS4B-4	The Pragmatic Functions of Prosody in English Cleft Sentences	
	<i>Sophie Herment, Laetitia Leonarduzzi</i>	713
PS4B-5	Prosody and discourse structure in French: a multilayer multi-parametric analysis of comic strip narrations	
	<i>Nassima Fezza, Cyril Auran</i>	717
PS4B-6	Context cues for classification of competitive and collaborative overlaps	
	<i>Catharine Oertel, Marcin Włodarczak, Alexey Tarasov, Nick Campbell, Petra Wagner</i>	721
PS4B-7	Prosodic Prominence and Phrasing in Spoken Mandarin: The Case of the 3rd tone	
	<i>Jianfen Cao</i>	725
PS4B-8	Proper-Name Identification	
	<i>Joost van de Weijer, Jeroen van de Weijer, Yan Beili</i>	729
PS4B-9	Response Types and The Prosody of Declaratives	
	<i>Catherine Lai</i>	733
PS4B-10	Question types and some prosodic correlates in 600 questions in the Spontal database of Swedish dialogues	
	<i>Jens Edlund, David House, Sofia Strömbergsson</i>	737
PS4B-11	Intonation Patterns of Wh-questions from Chinese EFL learners	
	<i>Ji Xiaoli, Wang Xia, Jia Yuan, Li Aijun</i>	741
K5	Bidirectional Tone Sandhi in Tianjin Dialect: Problem and Analysis	
	<i>Qiuwu MA</i>	745

Index

Analysis-by-Synthesis in Prosody Research

Rüdiger Hoffmann

Chair for System Theory and Speech Technology, Technische Universität Dresden, Germany

Ruediger.Hoffmann@tu-dresden.de

Abstract

It was early recognized in the history of speech technology, that prosody plays an essential role in the communication process and that it is therefore necessary to include prosodic components into the speech-based systems for man-computer interaction. Recent text-to-speech (TTS) systems show prosodic components at an elementary level (intonation and duration) for good comprehensibility, but it is also obvious that these components are not powerful enough to produce speech with high naturalness and personality. On the other hand, systems for automatic speech recognition (ASR) consider the prosody more or less implicitly, and we have only few examples where prosodic features are explicitly used for improving the recognition results. This talk is an attempt to give a more general view on the inclusion of prosody in speech technology. During the last decade, reconsidering the paradigm of analysis-by-synthesis (AbS) in speech technology has produced some algorithmic progress in TTS and in ASR as well. The system UASR (Unified Approach for Speech Synthesis and Recognition) of the TU Dresden was designed to demonstrate the AbS approach in a hierarchical way. It is now time to discuss how prosodic components could be included in such systems. The inclusion of rhythmic phenomena seems to be the most difficult but also very promising subtask. Possibly speech processing can benefit from musical signal processing where the identification of rhythm is a very natural task.

Index Terms: history of speech technology, Analysis-by-Synthesis, UASR, cognitive systems, hierarchical systems, speech dialogue systems

1. Introduction

Prosody research is growing very much during the last years. This is mainly due to the growing interest in social interaction where speech communication establishes only one of the communication modes. We have learned that speech prosody is not only part of linguistics, but also forms a bridge to non-linguistic communication and, above all, non-verbal modes like gesturing. Speech technology has utilized the progress in prosody research in a limited way until today. This is true for speech synthesis, but even more for speech recognition. Speech technology is now advancing towards speech dialogue systems. It seems to be useful, to reconsider the inclusion of prosody in such systems from an engineering point of view.

The investigation of prosodic effects in engineering has its own history. Roughly speaking, it started with a kind of trial and error, which was more and more refined to that epistemological approach which we call now Analysis-by-Synthesis (AbS).

AbS is very natural in speech technology because everybody will agree that building a speech based system means to design and implement a model of that what humans do if they are speaking or listening. AbS allows to optimize the modeling process (Figure 1) to achieve maximal similarity between the biological system and its engineering counterpart.

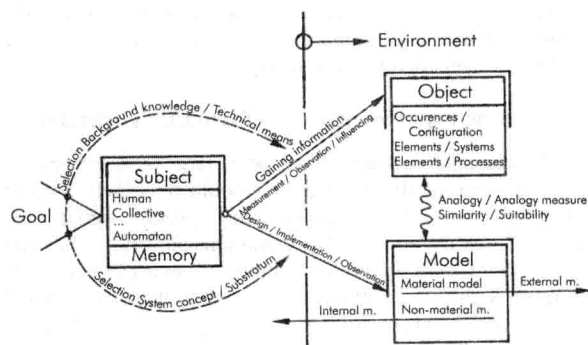


Figure 1: Model method (adapted from [1]).

2. AbS in the history of prosody research

2.1. The pre-electronic era [2]

It is interesting to note that Wolfgang von Kempelen, the forefather of the modern speech synthesis, recognized the importance of the speech melody for his speaking machine: "Ich habe oft nachgedacht, ob man nicht [...] dahin kommen könnte [...], dieses Fallen und Steigen des Tones nach Willkühr zu bewirken und dadurch [...] wenigstens eine Abwechslung der Stimme bey dem Sprechen zu erhalten, welches meiner Maschine, die dermalen alles in einem Tone fortspricht, erst die rechte Annehmlichkeit geben würde." [3, p. 413]. He describes first attempts with a manual control.

One century later, the special interest of the experimental phonetics in measuring the pitch contour as one of the most important physical phenomena of the prosody was activated because many foreign languages (the "colonial languages") had to be investigated. The analysis was performed mainly by interpreting the recordings of kymographs or phonographs. This very complicated and time-consuming process used a number of tools which we have described in [4]. Of course, there was no possibility to verify the results by means of re-synthesis.

2.2. The vocoder era

There were different attempts in speech synthesis at the beginning of the electronic era. The real breakthrough was achieved with the invention of the channel vocoder by K. O. Schmidt [5] and H. W. Dudley [6]. The subdivision of the device in an analyzer and a synthesizer enabled an analysis-by-synthesis process in a very effective way [7]. The existence of a separate channel for the fundamental frequency allowed the demonstration of the effect of pitch manipulation and thus the experimental investigation of prosodic contours. Some sound examples from the early vocoders are still available.

The analysis-by-synthesis activities in speech prosody go back to vocoder experiments. The linguists A. V. Isačenko (1910 - 1977, a well-known slavist) and H.-J. Schädlich

(* 1935, later known as a novelist) were among the first who developed models for the quantitative description of prosodic effects [8]. The English translation of their report [9] includes a disk with some of the test sentences. This test material consists of German sentences with a fundamental frequency which was manipulated to have only two values, e. g. [8]:

die Vorbereitungen sind getroffen, alles ist bereit

Experiments showed that there is still enough prosodic information to recognize the correct grammatical structure of the sentences. The manipulation was performed using the Dresden vocoder with support of W. Tscheschner and later with the Ericsson vocoder, supported by G. Fant.

2.3. Prosodic experiments with formant synthesizers

The first channel vocoders have been large and expensive. There was some doubt whether they could be widely used in commercial applications. Also, the speech signal had "inhuman" quality and limited comprehensibility. It became clear that there are more effective kinds of parameterization of the speech signal, and other vocoder types than the channel vocoder arose. Formant coding proved to be a very effective approach

Consequently, the early types of speech synthesis terminals also followed the principle of formant synthesis. This development was strongly influenced by the work of G. Fant and can be illustrated using the history at different places. We have described this way of early speech synthesis especially at the TU Dresden under the guidance of W. Tscheschner (1927 - 2004) in [10]. The prosodic investigations were connected to the ROSY project of the 1970-th. ROSY was a process computer controlled four-formant speech synthesizer. A small series of the synthesizers was produced by the Dresden computer company Robotron where the name of the device comes from (RObotron SYnthesizer). Formant synthesizers are very well suited for prosodic experiments (and even for singing) due to the presence of a separate excitation generator with controllable pitch.

The prosody research for the speech synthesizers of the TU Dresden was performed in close cooperation with the Humboldt University at Berlin. It can be divided in two phases. In the first one, the microintonation at the sound transitions of German was investigated using natural speech material. Different types of transitions were classified, and a group of five was finally proposed for the application in speech synthesis [11]. They were implemented in the hardware of the ROSY synthesizer.

In the second phase, analysis-by-synthesis experiments on the German macrointonation had been performed [12] with synthetic speech. For this purpose, the synthesis terminal ROSY was complemented by a contour generator which allowed influencing the intonation of the synthesizer by hardware. Basing on listening experiments, a number of standard contours could be proposed for the speech synthesis (Figure 2). Some examples of the test sentences in different intonation versions (monotonous / linear declination / declination plus accentuation) are still available as audio files.

2.4. Prosody in concatenative speech synthesis

The idea to synthesize natural sounding speech by concatenating speech segments from a database with real speech is not really new. With the invention of the magnetic storage of audio signals, the idea of the so-called concatenative synthesis emerged. The "digital" renaissance of the idea came with the availability of powerful PCs at the beginning of the 1990-th. They offered enough memory for the speech samples as well as enough computing power for the text and signal processing of the complete text-to-speech conversion chain. Unfortunately, prosodic manipulations were now more challenging compared to formant synthesizers. The TD-PSOLA algorithm [13] was the predominant solution and paved the way to a broad application of speech synthesis in time domain.

The emerging TTS technology required reliable control of the prosodic parameters for whole sentences or phrases. Therefore, quantitative models of macrointonation received more and more attention. A real breakthrough was achieved by the model of H. Fujisaki (e. g., [14]) which was applied successfully to many languages. Much effort was made to find effective training algorithms for the parameters of the Fujisaki model (e.g., [15]).

A systematical investigation of the German prosody was performed with the MFGI ("Mixdorff Fujisaki German Intonation") model. In this framework, we compared the prosodic quality of concatenative TTS for different prosody models and found that MFGI performed favorable [16].

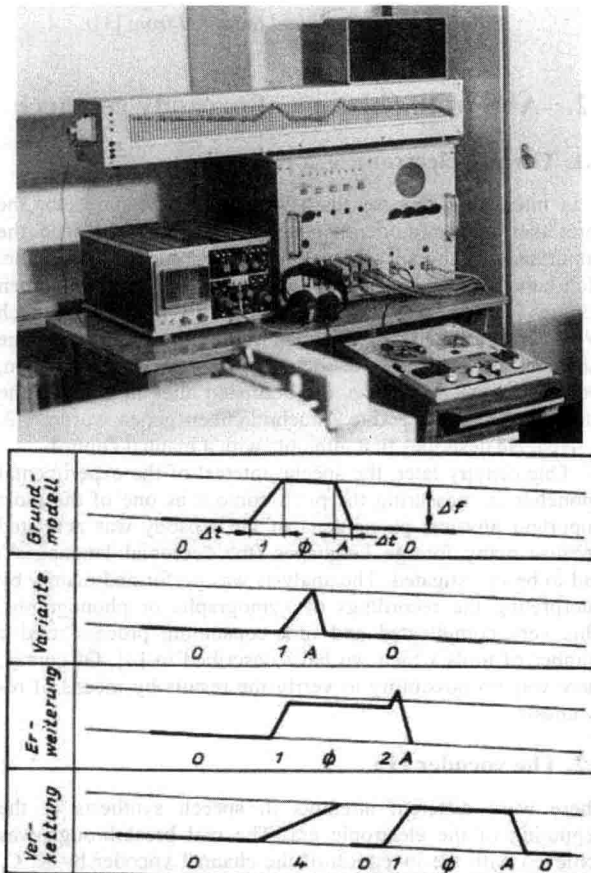


Figure 2 – Prosodic experiments with ROSY 4200. Above: Experimental setup with the synthesizer terminal ROSY (middle right) and the contour generator (above). The control computer is not shown. – Below: Models of suprasegmental fundamental frequency contours from [12].

3. From AbS to cognitive systems

3.1. The UASR platform as a prototype

The growing success of statistical approaches in speech technology during the 1990-th resulted in a convergence of speech recognition and speech synthesis which had developed hitherto in separate ways. This was mainly due to the necessity of large databases or knowledge sources in both branches. This development had been predicted in a classical textbook [17]: “Advanced systems both for synthesis and for recognition need the same speech knowledge, and there is considerable advantage for the two applications to be studied together. [...] I predict that the most significant progress in the more advanced forms of speech synthesis and recognition will in future come from research teams with a strong interest in both problems.” The development of the so-called HMM synthesis was the most important result of this generalized sight [18, 19].

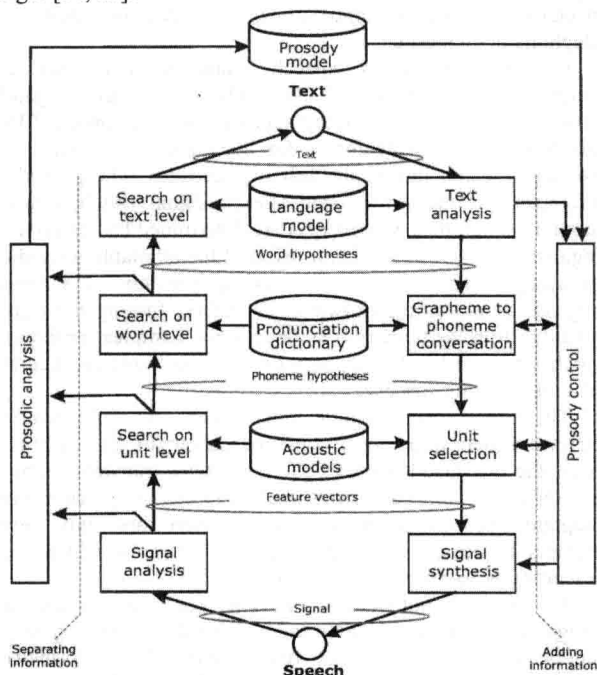


Figure 3: Unified approach for speech recognition and synthesis (UASR), as it was proposed in the year 2000 [20].

Recognizing these tendencies, we started around 2000 the development of a prototype system called UASR which integrated the elements of a typical speech recognizer and a speech synthesizer with common databases (Fig. 3). This system was implemented over the decade past 2000. About the progress, cf. [20 - 23]. The aims of the project have been:

- improved understanding of the algorithms by means of the principle of AbS in an *hierarchical* system,
- improved understanding of the reasons why speech recognition results are erroneous,
- development of components for parametric synthesis basing on statistically trained models,
- building a toolbox for practical (also embedded [24]) applications of speech recognition and synthesis,
- building the baseline system for numerous applications in the field of non-speech signals like biological [25], technical [26], or environmental signals, and music.

3.2. Cognitive dynamical systems

UASR is a prototype for a very up-to-date research field. S. Haykin coined the term *cognitive dynamic systems* for systems which show a purposeful behavior like human beings [27]. They are able to develop an internal model of their environment and, basing on this, to influence their environment actively. Obviously, there are close connections to the classical theory of automatic control (Fig. 4).

Surprisingly, elaborated applications of this theory are existing not only in the traditional fields of artificial intelligence (including speech technology), but also in “cognitive signal processing systems” like the cognitive radar [28] and the cognitive radio [29].

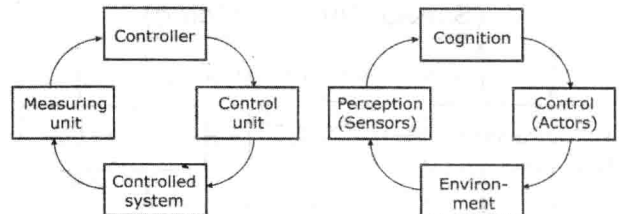


Figure 4: Cognitive dynamic system as proposed by Haykin (right), compared with the loop of a classical system for automatic control (left).

Considering the human as the most universal cognitive dynamic system, we must take into account the hierarchical structure of the information processing which is obviously essential for its function [30]. At all levels of the hierarchy, a combination of abstraction (bottom-up) and prediction (top-down) occurs. Therefore we find technical systems, which have (at least in a rudimentary way) a comparable hierarchical structure, mainly in the field of man-machine interaction, e. g. in processing speech, images, gestures, etc. This explains the formal analogy of the biological findings with the UASR structure in Fig. 3.

Other existing cognitive systems like cognitive radio require this hierarchical structure in less extent. A formal similarity exists due to the application of the well-known OSI reference model (Open Systems Interconnection Reference Model) which leads to the inclusion of the same statistical learning and decision algorithms.

Coming back to speech technology and AbS, it seems to be a natural extension of the UASR approach to form a cognitive system by adding a “speech understanding” component which acts as the cognitive module in the sense of Fig. 4. This could be very useful because it is generally recognized that the existing recognizers and synthesizers suffer from the lack of a real “understanding” of that what they do.

On a second glance, the problem arises how to design the interface between UASR and the understanding component. Speech understanding is a task of computer linguistics which normally expects an input of formally correct texts. This input, however, cannot be delivered by a system which processes *spoken* language due to two reasons. At first, natural (spontaneous) speech is not regular in a strong sense. Secondly, we know that a speech recognizer makes errors.

These problems had been already considered in a former big research project, called Verbmobil, where the understanding component was a translation software [31]. We can apply experiences from this project if we are enlarging the UASR structure.