

Session 3aSCa

Speech Communication: Psychoacoustic Methods in Speech Recognition I

Birger Kollmeier, Cochair

AG Medizinische Physik, Universität Oldenburg, Postfach 2503, D-26111 Oldenburg, Germany

Hynek Hermansky, Cochair

Oregon Graduate Institute of Science and Technology, Beaverton, Oregon 97006-8921, USA

Chair's Introduction—7:55

Invited Papers

10:20

3aSCa5. Syllable-based speech recognition using auditorylike features. Steven Greenberg, Takayuki Arai, Brian Kingsbury, Nelson Morgan, Michael Shire, Rosaria Silipo, and Su-Lin Wu (International Computer Science Inst., 1947 Center St., Berkeley, CA 94704)

Classical models of speech recognition (both human and machine) assume that a detailed, short-term analysis of the signal is essential for accurate decoding of spoken language via a linear sequence of phonetic segments. This classical framework is incommensurate with quantitative acoustic/phonetic analyses of spontaneous discourse (e.g., the Switchboard corpus for American English). Such analyses indicate that the syllable, rather than the phone, is likely to serve as the representational interface between sound and meaning, providing a relatively stable representation of lexically relevant information across a wide range of speaking and acoustic conditions. The auditory basis of this syllabic representation appears to be derived from the low-frequency (2–16 Hz) modulation spectrum, whose temporal properties correspond closely to the distribution of syllabic durations observed in spontaneous speech. Perceptual experiments confirm the importance of the modulation spectrum for understanding spoken language and demonstrate that

1157 J. Acoust. Soc. Am., Vol. 105, No. 2, Pt. 2, February 1999

Joint Meeting: ASA/EAA/DEGA

1157

the intelligibility of speech is derived from both the amplitude and phase components of this spectral representation. Syllable-based automatic speech recognition systems, currently under development, have proven useful under various acoustic conditions representative of the real world (such as reverberation and background noise) when used in conjunction with more traditional, phone-based recognition systems.

1158 J. Acoust. Soc. Am., Vol. 105, No. 2, Pt. 2, February 1999

Joint Meeting: ASA/EAA/DEGA

1158