50. Jahrestagung
25.- 29.9.2000 Karl-Franzens Universität Graz
Physik-Event ÖPG 2000

Jahrestagung der ÖPG
FA-Akustik

gemeinsam mit
der Österreichischen Akustischen Gesellschaft (AAA)

Auditory time-frequency analyses applied to phoneme segmentation

Phonemes, vowels and consonants are defined linguistically and represented acoustically in a speech sequence. However, phonemes might also be explained by their psychoacoustic relevance. This study compares the presence of vowel boundaries within the waveform between standard time-frequency analyses and analyses based on auditory filter models. Both auditory imaging based on a gammatone filterbank (2,6) and LPC-formant estimation based on short-term Fourier transformation do not resolve clear vo wel boundaries.

In order to implement automatic segmentation by hidden Markov models, a pitch-synchronous auditory analyses is proposed. Voiced speech is generally modeled as a sequence of glottal pulses filtered by the vocal tract (1). A speech sequence can be interpreted as a series of impulse responses of the vocal tract filter. In the analysis each speech period was extracted by a peak-search algorithm, containing a speaker specific decision process. Each period was padded to a sequency and frequency-domain filtered by a set of auditory ROEX-filters (4), with bandwidths based on the ERB-scale (3). The position of peaks and valleys of the resulting auditory representation are compared to the peaks and valleys obtained by LPC-formant analysis for three sets of stressed German vowels of read speech (/a/,/I/,/aI/). Both methods result in temporal edges within the first three periods.

References

[1] Fant, G. Analysis and synthesis of speech processen, Manual of Phonetics; Bertil Malmberg Ed.; North-Holland Publishing Co. Amsterdam 1974; p.173-277.
[2] Ghitza O., Adequacy of auditory models to predict human internal representation of speech sounds. J Acoust Soc Am. 1993 Apr;93(4 Pt 1):2160-71.
[3] Glasberg BR, Moore BC., Derivation of auditory filter shapes from notched-noise data. Hear Res. 1990 Aug 4; 47(1-2):103-38.
[4] Patterson RD. Auditory filter shapes derived with noise stimuli. J Acoust Soc Am. 1976 Mar;59(3):640-54.
[5] Plack CJ, Moore BC. Temporal window shape as a function of frequency and level. J Acoust Soc Am. 1990 May;87(5):2178-87.
[6] Tchorz J, Kollmeier B., A model of auditory perception as front end for automatic speech recognition. J Acoust Soc Am. 1999 Oct;106(4 Pt 1):2040-50.

Dipl.phys Bram G. Alefs
Austrian Academy of Sciences - Acoustics Research
A-1010 Vienna, Austria
Tel. 431 4277 29501 Fax 431 4277 9296 - Email: bram@kfs.oeaw.ac.at - http://www.kfs.oeaw.ac.at