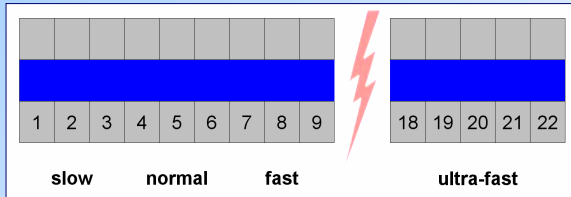


FAST SPEECH

Speech rate classification



Fast speech characteristics:

Shortening, reduction, assimilation and elision of phones and syllables

- cause a worse perception of natural fast speech,
- are not applicable to produce fast speech in a unit selection synthesis system,
- therefore have to be avoided.

A speaker was selected

- who was a skilled speaker,
- who was able to speak both very fast and clear.

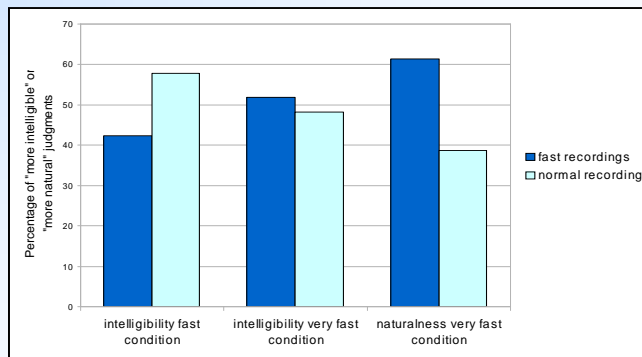
400 sentences were recorded in 2 conditions:

- normal,
- fast and clear.

EVALUATION

Linear acceleration by TD-PSOLA in 2 steps:

- normal rate sentences accelerated to tempo of fast rate sentences,
- normal as well as fast rate sentences accelerated to very fast rate (double fast rate).



Fast condition: normal rate stimuli judged to be more intelligible than natural fast spoken ones.

Very fast condition: fast rate stimuli clearly preferred with respect to naturalness.

DURATION PREDICTION (CART)

Speech rate	Correlation	Mean (absolute) Error (ms)
Normal	0.82	20.16 (34.16)
Fast	0.78	12.37 (20.53)

No significant differences between normal and fast speech rate for the correlation between observed and predicted segment duration.

No significant differences in feature ranking between normal and fast speech.

REFERENCES

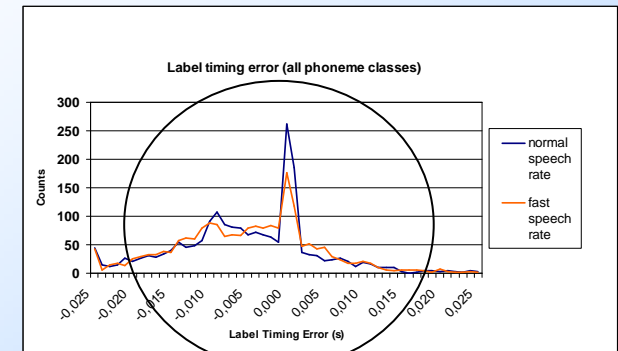
Moers, D., Wagner, P. and Breuer, S., "Assessing the Adequate Treatment of Fast Speech in Unit Selection Speech Synthesis Systems for the Visually Impaired", *Proc. 6th ISCA Workshop on Speech Synthesis (SSW-6)*, Bonn, 2007.
 Moers, D. and Wagner, P., "Assessing a Speaker for Fast Speech in Unit Selection Speech Synthesis", *Proc. Interspeech 2009*, Brighton, 2009.
 Moos, A. and Trouvain, J., "Comprehension of Ultra-Fast Speech – Blind vs. "Normally Hearing" Persons", in: *Proceedings ICPhS XVI*: 677-684, Saarbrücken, 2007.
 Kohler, K.J., "Segmental reduction in connected speech in German: Phonological facts and phonetic explanations", in: Hardcastle, W.J. and Marchal, A. [Ed.], *Speech Production and Speech Modeling*: 69-92, Dordrecht, 1990.
 van Son, R. J. J. H. and Pois, L. C. W., "An acoustic profile of consonant reduction", in *Proceedings ICSP*: 1529-1532, Philadelphia, 1996.
 Crystal, T.H. and House, A.S., "Articulation rate and the duration of syllables and stress groups in connected speech", *Journal of the Acoustical Society of America* 88: 101-112, 1990.
 Monaghan, A., "An Auditory Analysis of the Prosody of Fast and Slow Speech Styles in English, Dutch and German", in Keller, E. et al. [Ed.], *Improvements in Speech Synthesis*, 204-217, Chichester, 2001.
 Janse, E., "Word perception in natural-fast and artificially timecompressed speech", *Proceedings ICPhS XV*: 3001-3004, Barcelona, 2005.
 Schiel, F. et al., "Die BITS Sprachsynthesekorpora – Diphon- und Unit Selection-Synthesekorpora für das Deutsche", *Proceedings Konvens 2006*: 121-124, Konstanz, 2006.

Feature	Correlation Normal	Correlation Fast
Phone identity	0.47	0.47
Phrase position	0.68	0.67
Follow. Phoneme	0.79	0.76
Preceding phoneme	0.80	0.77
Syllabic stress	0.81	0.77
2nd follow. phoneme	0.82	0.78

AUTOMATIC ALIGNMENT

49 sentences of each corpus labeled manually.

Label timing differences calculated by subtracting manual label time from automatic label time.



No significant differences between normal and fast speech rate for a 20 ms tolerance interval.

Speech rate	< 5 ms	< 10 ms	< 15 ms	< 20 ms
Normal	42.8%	68.8%	83.0%	90.4%
Fast	42.6%	65.7%	82.0%	90.8%