

Evaluation eines Sprechers für schnell gesprochene Sprache in der Unit-Selection basierten Sprachsynthese

Donata Moers, Petra Wagner

Institut für Kommunikationswissenschaften, Abt. Sprache und Kommunikation,
Rheinische Friedrich-Wilhelms-Universität Bonn
{dmo,pwa}@ifk.uni-bonn.de



universität**bonn**

EINLEITUNG - INTRODUCTION



Wie erfassen Sie, ob dieses Bild Informationen enthält, die für Sie interessant sein könnten?

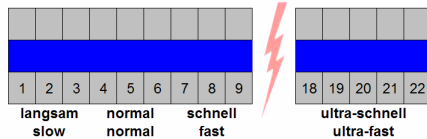
How do you capture if this picture contains information which could be interesting for you?

MOTIVATION

- ▶ Blinde und Sehbehinderte bevorzugen bisher die Formantsynthese: Zufall? Mangel an Alternativen?
Up to now, the blind and visually impaired prefer formant synthesis: Hazard? Little alternatives?
- ▶ Ist mit konkatenerativer Sprachsynthese eine höhere Natürlichkeit und größere Akzeptabilität zu erreichen?
More naturalness and higher acceptability possible with concatenative synthesis?
- ▶ Ist die Realisierung schneller Sprache in der konkatenerativen Synthese überhaupt möglich?
Realization of fast speech in concatenative synthesis possible at all?

SCHNELLE SPRACHE - FAST SPEECH

Silben pro Sekunde – syllables per second



- ▶ Verkürzung, Reduktion, Assimilation, Elision von Lauten und Silben
Shortening, reduction, assimilation, elision of phonemes and syllables
- ▶ Weniger Intensität, weniger Intonation, weniger Phrasen, weniger Pausen
Less intensity, less intonation, fewer phrases, fewer pauses

Bei der Modellierung schneller Sprache in der Unit-Selection basierten Sprachsynthese sind diese Phänomene größtenteils unerwünscht, da sie nicht nur die Verständlichkeit der erzeugten Sprache negativ beeinflussen, sondern auch die Definition der Auswahlseinheiten wesentlich erschweren.

Many of this characteristics cause an even worse perception of natural fast speech. So this are all aspects of fast speech which are not applicable to produce fast speech in a unit selection synthesis system and therefore have to be avoided.

SPRECHSTRATEGIEN - SPEAKERS' STRATEGIES

▶ H&H-Theorie

1. Ökonomie: Möglichst geringer Artikulationsaufwand → Hypospeech
2. Zweckorientiertheit: Kommunikationsziel erreichen mittels phonetischer Kontrastierung → Hyperspeech

Es ist zu erwarten, dass ein Sprecher sich ständig entlang des Kontinuums zwischen Hypo- und Hyperspeech bewegt. Bei schnellem Sprechen erwarten wir normalerweise Hypospeech – aufgrund der ökonomischen Voraussetzungen. Dennoch müssen Sprecher innerhalb der artikulatorischen Grenzen in der Lage sein, schnell und deutlich (Hyperspeech) zu sprechen, wenn die Kommunikationssituation dies erfordert.

▶ H&H-theory

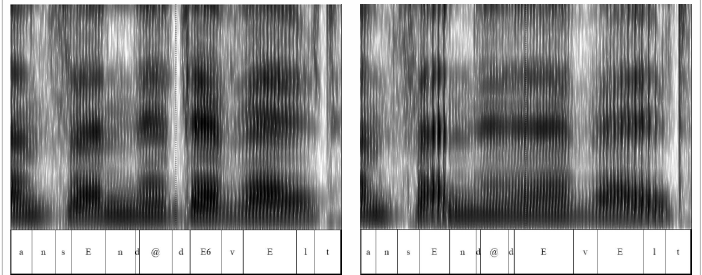
1. Economics: As little articulatory effort as possible → Hypospeech
2. Goal oriented: Reach aim of communication by maintaining phonetic contrast → Hyperspeech

Speakers are expected to vary their output along a continuum of hyper- and hypospeech to reach a communicative goal. For fast speech, we would normally expect speakers to use hypospeech while speaking fast – due to economy. However, speakers may be well able to speak both fast and clear (Hyperspeech) if the situation requires this – within certain articulatory constraints.

VORAUSSWAHL - PRESELECTION

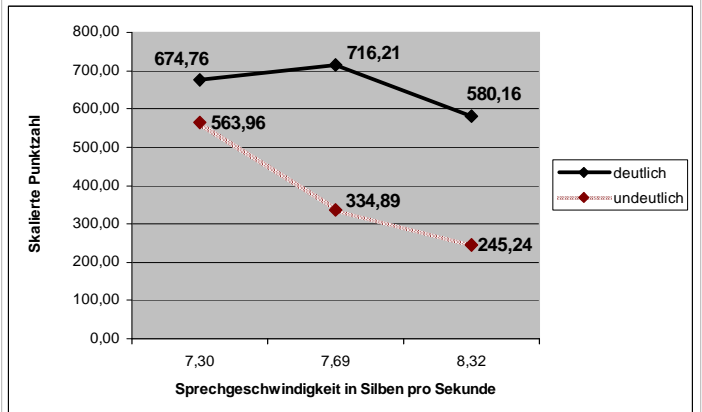
- ▶ 9 freiwillige Sprecher mit Sprecherfahrung aus verschiedenen Bereichen
9 voluntary experienced speakers
- ▶ Aufnahmen in normalem und schnellem Sprechtempo
Recordings in normal and fast speech tempo
- ▶ Perzeptive Beurteilung durch geschulte Hörer
Perceptive evaluation by experienced listeners
- ▶ Weitere Kriterien
Other criteria

EVALUATION



- ▶ 3 schnelle deutliche und 3 schnelle undeutliche Versionen
3 fast and clear as well as 3 fast and inarticulate versions
- ▶ Perzeptionstest mit geschulten und ungeschulten Hörern; Paarvergleich aller Versionen
Perception experiment with experienced and inexperienced listeners; paired comparison of all versions
- ▶ Normierung der erreichten Punktzahl sowie Umrechnung der Punktzahl auf die jeweilige Sprechgeschwindigkeit
Standardization of achieved scores as well as mapping scores onto respective speech tempo

ERGEBNISSE - RESULTS



AUSBLICK - PROSPECT

Da sich die ausgewählte Sprecherin tatsächlich als fähig erwiesen hat, extrem schnell bei immer noch sehr deutlicher Artikulation zu sprechen, wird als nächstes ein größeres Korpus in schnell gesprochener Sprache aufgenommen werden, welches anschließend als Bausteininventar für die Erzeugung synthetischer schnell gesprochener Sprache in unserem Unit-Selection basierten Sprachsynthesensystem BOSS verwendet und im Vergleich zu einem normal gesprochenen Syntheseinventar weiter evaluiert werden wird.

The chosen speaker actually is qualified for the task to speak extremely fast but still clear at the same time. Therefore, the next step will be the recording of a more substantial corpus of fast speech which will serve as inventory for the generation of fast synthetic speech using our unit selection based synthesis system BOSS. The fast speech inventory will be further evaluated by comparing fast synthetic speech produced from a normal tempo inventory versus the fast speech inventory.

LITERATUR - REFERENCES

- Breuer, S.; Abresch, J. (2004): Phoxsy: Multi-phone Segments for Unit Selection Speech Synthesis. In Proceedings ICSLP, Jeju.
- Crystal, T.H.; House, A.S. (1990): Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, Vol. 88, S. 101 – 112.
- Goldman-Eisler, F. (1961): The significance of changes in the rate of articulation. *Language and Speech*, Vol. 4, S. 171 – 174.
- Gopal, H.S. (1990): Effects of speaking rate on the behaviour of tense and lax vowel durations. *Journal of Phonetics*, Vol. 18, S. 497 – 518.
- Greisbach, R. (1992): Reading aloud at maximal speed. *Speech Communication*, Vol. 11, S. 469 – 473.
- Janse, E. (2003): Production and Perception of Fast Speech. *Dissertation*, Universiteit Utrecht.
- Kohler, K.J. (1990): Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In *Hardcastle, W.J.; Marchal, A. (eds.): Speech Production and Speech Modelling*. Dordrecht: Kluwer, S. 69 – 92.
- Lindblom, B. (1990): Explaining phonetic variation: A sketch of the H&H-Theory. In *Hardcastle, W.J.; Marchal, A.: Speech Production and Speech Modelling*, Dordrecht: Kluwer, S. 403 – 439.
- Moers, D.; Wagner, P.; Breuer, S. (2007): Assessing the Adequate Treatment of Fast Speech in Unit Selection Speech Synthesis Systems for the Visually Impaired. In *Proceedings 6th ISCA Workshop on Speech Synthesis (SSW-6)*, Bonn.
- Monaghan, A. (2001): An Auditory Analysis of the Prosody of Fast and Slow Speech Styles in English, Dutch and German. In *Keller, E.; Bailey, G.; Monaghan, A. et al. (eds.): Improvements in Speech Synthesis*. Chichester, S. 204 – 217.
- Moos, A.; Trouvain, J. (2007): Comprehension of Ultra-Fast Speech – Blind vs. ‘Normally Hearing’ Persons. In *Proceedings ICPHS XVI*, Saarbrücken, S. 677 – 684.
- van Son, R. J. J. H.; Pois, L. C. W. (1996): An acoustic profile of consonant reduction. In *Proceedings ICSLP, Philadelphia*, S. 1529 – 1532.