Speech Synthesis of Dialectal Variants as a Method for Research on Prosody

Beat Siebenhaar, Martin Forst, Eric Keller

With thanks to Brigitte Zellner Keller

Laboratoire d'Analyse Informatique de la Parole

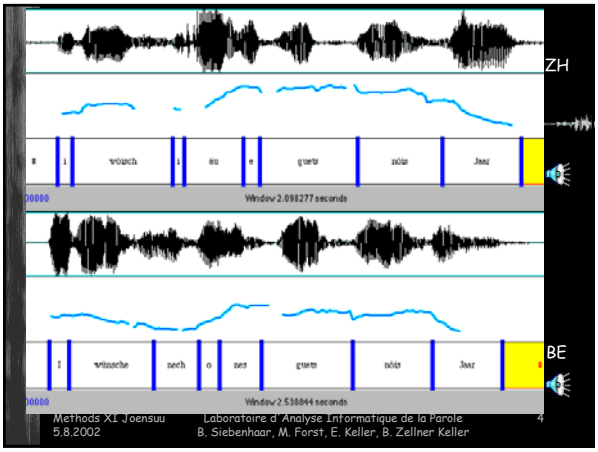
Université de Lausanne

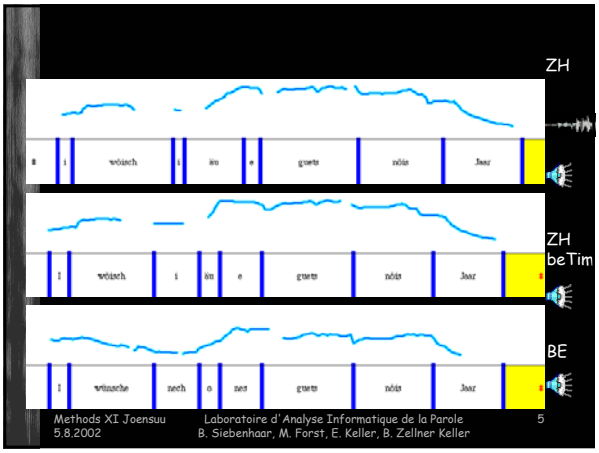
Introduction

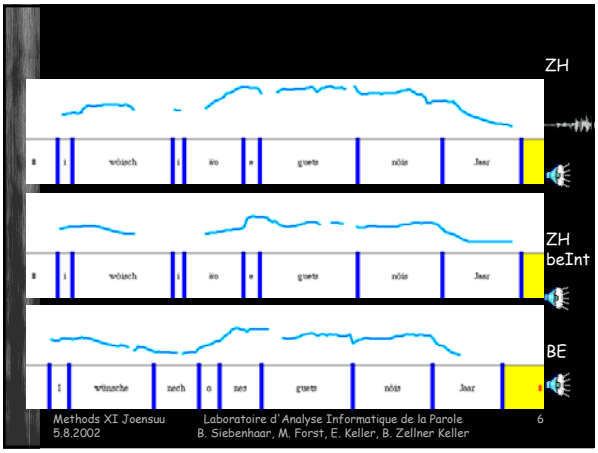
- Speech synthesis has reached quality levels that henceforth permit its use in testing linguistic hypotheses. (Keller, Keller Zellner 2000)
- Outline
  - Problematic Prosody Modification
  - Speech Synthesis as a Tool for Linguistic Research
  - Architecture and History of LAIPTTS (LAIP Text-to-Speech Synthesis)
  - Speech Synthesis for Dialectal Variants
  - Prosody Modification by a Dialectal Synthesis

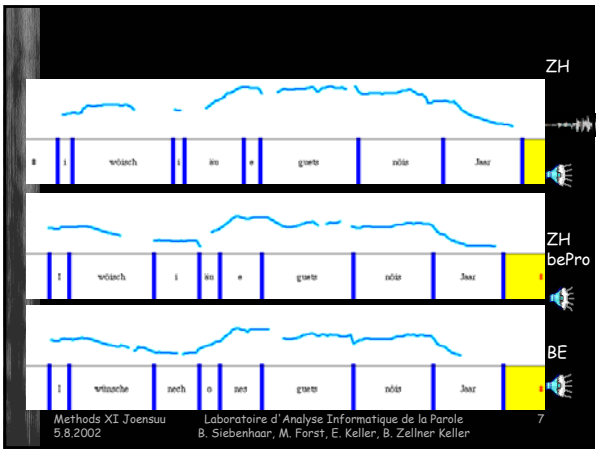
Problematic Prosody Modification

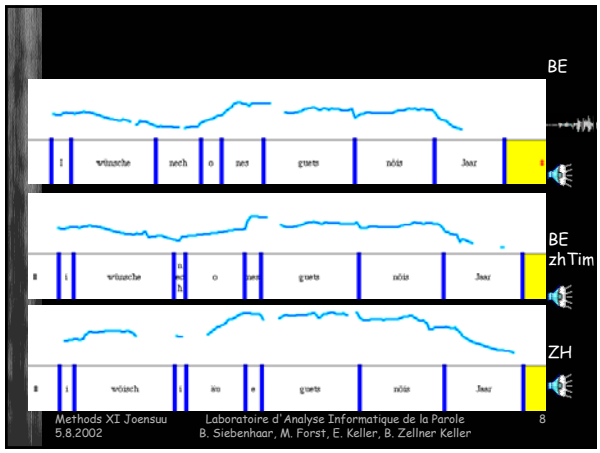
- Matched guise techniques are used to exclude speaker-specific ratings of speech signals.
- Prosodic parameters are very difficult to control.
- Researches rely on manipulating the original speech signals. What we are allowed to change to obey the prosodic grammar of a language?

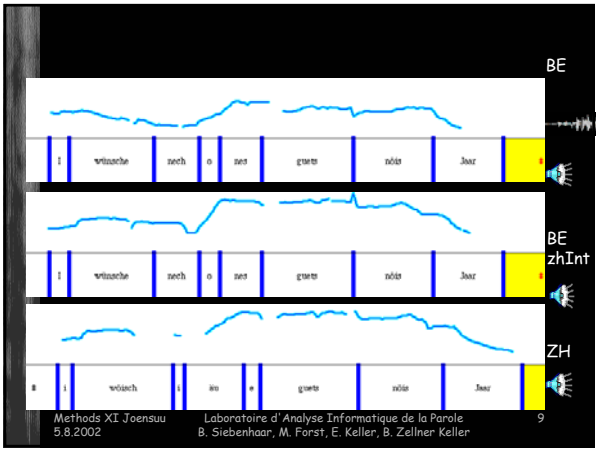


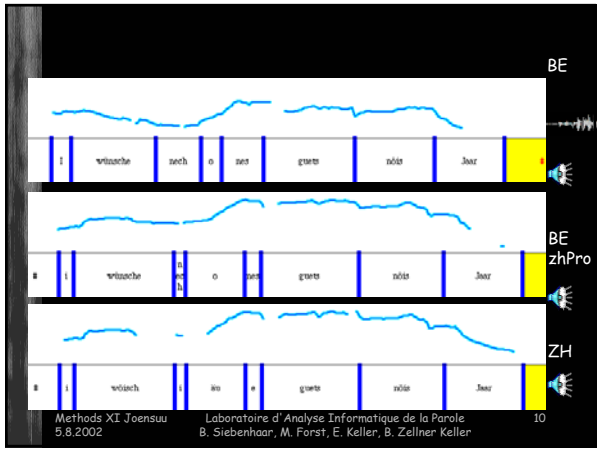












Problematic Prosody Modification

- Even with a prosody transplantation of almost identical texts, often the result is not satisfying.
- It is problematic to make prosodic changes manually in a consistent way.
- As soon as one wants to change syntactic and morphologic aspects, this task is almost impossible.

Speech Synthesis as Tool for Linguistic Research

- The traditional scientific method is **analysing** data and explaining the findings about a certain phenomenon. Apparently non-relevant phenomena are excluded.
- **Synthesis** as a scientific method **forces** us to model interactions between the different subsystems.

Speech Synthesis as Tool for Linguistic Research

- Utterances with wrong models for phrasing and timing.

•Imagine how difficult it would be to understand a speech signal where phrase boundaries are placed in an arbitrary fashion.
Stellen Sie sich vor, wie schwer es wäre ein Sprachsignal zu verstehen, wo Phrasengrenzen vollkommen willkürlich gesetzt werden.

Imagine a speech signal where the underlying timing model does not take phrase boundaries into account. You hear that the sounds before pauses are too short, because among other things the final lengthening is not respected. Pauses are therefore badly perceived; the utterance sounds choppy.
Stellen Sie sich ein Sprachsignal vor, in dem das zugrundeliegende Modell für die zeitliche Steuerung keine Phrasengrenzen berücksichtigt. Sie hören, dass die Laute vor den Pausen zu kurz sind, weil unter anderem keine Längung am Phrasenende eingebaut ist. Pausen werden so relativ schlecht wahrgenommen; die Sprache wirkt abgehackt.

Speech Synthesis as Tool for Linguistic Research

- For a speech synthesis system, we need an adequate model for every aspect of prosody. One deficient model can obscure the adequate models of other aspects in the resulting speech output.
- With a synthesis approach we might gain new insights complementary to the information one can obtain with an analytic approach.
- Speech synthesis can be a diagnostic tool.
- Synthesis forces us to elaborate global aspects first, which then can be refined step by step.

Architecture and History of LAIPTTS

- The design of our dialectal synthesis system will follow the "LAIP tradition" of speech synthesis concerning major theoretical decisions
- History of LAIPTTS
  - Development of a psycholinguistically motivated prosody model for French → LAIPTTS_F (Brigitte Zellner Keller & Eric Keller)
  - Adaptation of this model to Standard German → LAIPTTS_D (Beat Siebenhaar & Martin Forst)
  - Our project: Further adaptation of the model to Bernese and Zurich German (Beat Siebenhaar & Martin Forst)

Architecture of LAIPTTS

- Linear model including (in this order)
  - a robust word grouping algorithm, based mainly on the distinction of grammatical and lexical words
  - a component for the calculation of segment durations, consisting of a general linear model
  - an intonation component based on the Fujisaki approach of F0 calculation
- The model can thus represent influences of phrasing on timing and intonation and repercussions of timing on intonation, but not vice versa.

Architecture of LAIPTTS

- Further information about our TTS system can be found at www.unil.ch/imm/docs/LAIP/LAIPTTS.html

Architecture of LAIPTTS: the word grouping algorithm

- Main rule: Minor phrase boundaries are inserted after every lexical word followed by a grammatical word
- Additional rules for special contexts (fixed expressions, negations, complex verbal expressions, etc. (Zellner, 1996))
- Major phrase boundaries are inserted at punctuation marks. Then it is checked if the number of syllables between major breaks exceeds a certain threshold. As long as that is the case, the minor phrase boundaries closest to the middle of such major prosodic phrases are upgraded to major break status. As a result we obtain evenly balanced prosodic phrases.

Architecture of LAIPTTS: calculation of segment durations

- Statistic (general linear) models built on the data of manually labelled corpora
- Parameters taken into account :
  - the durational class of the current segment
  - surrounding segments
  - the structure of the syllable the segment belongs to
  - the grammatical status of the corresponding word
  - the position of the segment in the word
  - the position of the segment in the prosodic phrase

Architecture of LAIPTTS: the intonation component

- Superpositional Fujisaki model (Fujisaki, Hirose 1982):
  - relatively slow phrase commands that determine the general intonation contour in a prosodic phrase
  - relatively fast accent commands
  - The resulting curves of both kinds of commands are added and result in the final F0 contour.
- The position, duration and slope of these commands have been obtained from a copy-synthesis-analysis and statistical classification of the respective parameters.

Architecture of LAIPTTS: Major Changes for LAIPTTS-D

- Word grouping algorithm: insertion of a minor break in front of the verb in main clauses
- Calculation of segment durations: Lexical stress needs to be taken into account
- Intonation component: Some changes which are also due to lexical stress

Speech Synthesis for Dialectal Variants

- Major problems to resolve before elaborating the model:
  - The models for French and standard German are based upon an analysis of read speech, reproducing a news reading style considered as neutral.
  - The Swiss German dialects are hardly written, nor are they the language of the news usually. So we cannot model the typical TTS style.
  - → Not only new languages/dialects, but also a different style of speech, probably with more intrinsic variation on the prosodic level.

Speech Synthesis for Dialectal Variants

- Our choice: Interview style
  - Naturally embedded into a communication situation
  - Certain degree of formality
  - Combines best naturalness and control of the language
  - work on different speech rates (Zellner 1998) and pre-tests in different speaking styles have shown that phrase boundaries are very similar across varieties
  - LAIPTTS word grouping algorithm is robust enough to work reasonably well with ungrammatical input

→ The general approach can be applied to the new data.

Speech Synthesis for Dialectal Variants

- Advantage of choosing the same approach for TTS systems of different languages, dialects and styles:
  - same or similar parameters taken into account → direct comparison of models
- Particularity of our project:
  - Bernese and Zurich German are closely related, yet clearly different. It is possible to elaborate parallel prosodic models
  - Differences of parameters inform about the way in which prosody can vary independently of the segmental information

Outlook to Solutions by a Dialectal Synthesis

- In a speech synthesis system, each prosodic parameter can be modified independently of the others while, at the same time, the interaction of the different aspects of prosody being maintained. So sound examples can be generated with different models at each linguistic level. In perception tests, we can evaluate the importance of the different models taken one by one
- For instance, a text with lexical, syntactical and morpho-logical information characteristic for the Zurich dialect can be synthesized with the Bernese prosodic model, or with a Bernese timing model, but a Zurich model for intonation.

Outlook to Solutions by a Dialectal Synthesis

- Speech synthesis system lets one manipulate the different prosodic parameters consistently in the same fashion.
- Example: This is a Frech text, prodced with the rules for German rythm and intonation, but generated with a French voice.
  Voici un texte français qui est produit avec les règles pour le rythme et l'intonation de l'allemand mais avec la voix française.
- *French prosodics*
- *German prosodics*
- French *prosodics*: voici…
- German *prosodics*: voici…

Prosody Modification by a Dialectal Synthesis

- Speech synthesis can be a way out of the dilemma of manipulating the related aspects of prosody.
- With this synthesis tool we hope to obtain answers about the general structure of the prosody of the two dialects. And while building the system, we will no doubt detect the black holes in our knowledge of prosody.
- At the end we will be able to generate sound examples with different prosodic models that can be submitted to perception tests that may provide hints about the importance of different phonetic and prosodic aspects in dialect classification by naive listeners.

Thank you

Thank you for visiting us at:
http://www.unil.ch/imm/docs/LAIP/LAIP.html