A4

Phonetic Convergence in Spontaneous Speech

SFB colloquium
Blaubeuren
November 14-16, 2008
Overview

- What is phonetic convergence?
- overview of previous studies
- Natalie's dissertation project on convergence in non-native speakers
- data processing for examining convergence
- Which data are appropriate?
- Which measures are appropriate?
Phonetic Convergence

<table>
<thead>
<tr>
<th>Phonetic Convergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>convergence: adaptation to the communicational behaviour of an interlocutor</td>
</tr>
<tr>
<td>opposite: divergence</td>
</tr>
<tr>
<td>may occur on various linguistic levels</td>
</tr>
<tr>
<td>choice of language in bilinguals, use of accent or dialect</td>
</tr>
<tr>
<td>information density and self-disclosure</td>
</tr>
<tr>
<td>semantic and conceptual structure</td>
</tr>
<tr>
<td>syntactic constructions</td>
</tr>
<tr>
<td>prosody (pitch range, timing)</td>
</tr>
<tr>
<td>segmental aspects (pronunciation)</td>
</tr>
</tbody>
</table>

Introduction/Previous studies
Convergence in non-native speakers
Data processing and data for A4 Experiments
Phonetic Convergence

- unconscious and dynamic process according to Communication Accommodation Theory (CAT, Giles et al. 1979-2006) and Alignment Theory (Pickering & Garrod 2004)
- modulates social distance according to CAT, social roles may influence readiness to converge
- like exemplar theory, assumes link between perception and production
- demonstrates dynamic aspect of exemplar updates and decay
- dynamic nature of effect requires incremental investigation of fine phonetic detail
Previous Studies

- Matarazzo & Wiens 1967, Matarazzo et al. 1963
  - job interviews, 20 subjects in each experiment
  - interviewer **systematically varied** response latency RL and utterance duration UD (naturalness?) in three 15-minute blocks
  - mean RL: subtle, not significant for most subjects
  - mean UD: significant for most subjects
Previous Studies

• Natale 1975
  – similar to Matarazzo, but three 10-minute blocks
  – interviews (21) and conversations (50) via loudspeaker
  – vocal intensity of the interviewer/conversation partner was
    manipulated by limiting amplifiers
  – subjects converged significantly with respect to averaged
    intensity
  – **degree of convergence predicted by a "social desirability score"** measuring each subject's conformity
Previous Studies

• Cappella & Planalp 1981
  – *time-series regression to capture dynamic aspects* of the conversation
  – past parameters of both speakers predict current parameters
  – speakers' own parameters have more impact than partners' parameters (self-consistency)
  – *both convergence and divergence* observed!
Previous Studies

• Street 1984
  – interviews of students with business persons
  – mutual ratings of social attractiveness and competence afterwards
  – time series regression revealed convergence with respect to speech rate and response latency
  – both convergence and divergence for turn duration
  – degree of convergence correlated with ratings
Previous Studies

  – 130 interviews conducted by the author 10 years before
  – each publication examines just few randomly picked interviews
  – **long-term average spectra** (LTAS) of participants are correlated
  – conversation pairings can be predicted by LTAS similarities
Previous Studies

• Gregory & Webster 1996, Gregory & Gallagher 2002
  – **factor analysis of LTAS** in Larry King talk show and presidential candidates' TV debates
  – **factor loadings correspond to social factors**
    • less prominent guests converge to King; King converges to more prominent guests
    • successful candidates converge less
Previous Studies

• Pardo 2006
  – **map task experiment**: 6 pairs of interlocutors
  – participants read map task landmarks before and after the experiment (read speech)
  – **perceptual ratings of similarity** of the landmarks' pronunciation by independent subjects
  – ratings increased from before the map task to early in the map task to later in the map task and after the map task -> convergence
  – effects of role and gender
Lewandowski 2008: Convergence in non-native dialogs

- increase in match values between early and late point in the dialogs for both conditions indicating phonetic convergence
- AE condition: significant increase in match values for the talented speakers, no significant increase for the less talented speakers
- BE condition: no significant differences
- no significant gender effects
- no significant effect for the preferred accent
- no generalization over speech varieties (read spoken language)
Diapix - dialog-based picture matching (Bradlow et al. 2007)

• “spot-the-difference” game involving 2 pictures and 2 participants.
• Participants cannot see each other’s picture.
• They work together to find 10 differences (3 missing from each version & 4 changed items).
• Elicits a wide range of utterance types (questions, declaratives, exclamations etc.)

• balanced talker roles
• task-oriented but spontaneous speech
• conversations last approx. 10 to 20 minutes
Examples of convergence & match values

- bullets A X B  A=0,9013
- carrots A X B  B=0,6836
- wline A X B  A=0,8875
- house A X B  B=0,8357
IMS previous work for generating reliably annotated large corpora

- Manual transliteration possibly using ASR
- German morphology (SmartWeb project)
- POS Tagger (Schmid 1995)
- Aligner (Rapp 1995)

- Speech signal
- PaintE
- Parametrization of F0 (Möhler & Conkie 1998)
- Automatic annotation of GToBI (Schweitzer, work in progress)

- Prosodic transcription
- Mostly automatic annotation allows for much larger reliably annotated corpora
Predicting prosodic events

- idea: predict prosodic events based on PaIntE parametrization and zscores of durations, but also use text-based linguistic information
- hypothesis: humans do not perceive prosody without linguistic context
- corpus: speech synthesis corpus from male speaker, 2 1/2 hours, professional speaker, manual prosodic annotation
- WEKA to train classifiers for prediction
Predicting prosodic boundaries

- classes predicted: (H)-, (L)-, H%/H)%, (L)%, NONE
- best algorithm: random forests
- correctly classified syllables: about 95%, corresponds to more than 90% for words
- about 92% on word level for classification of break/no break
Predicting pitch accents

- best algorithm: random forests
- correctly classified syllables about 87%, corresponds to ?% for stressed syllables only
- actually doesn't predict ..L, ..H on test data
Predicting prosodic events - outlook

• forests generalize to female data (but partly identical text material)
• Festival can use the forests to generate prosodic label files for whole databases
Possible data

- conversations from radio talk shows
  - acquisition easy, large amounts of data available, excellent recording quality, professional speakers
  - repeated recordings of one speaker with different guests
  - problem: social relations must be rated afterwards
- own diapix recordings
  - manipulation of social roles possible
  - subjects available for assessment of attitude afterwards
  - fully controlled: status, topics of the conversation
  - acquisition time-consuming, non-professional speakers
Experiments

- segmental convergence
  - amplitude envelopes (A2)
  - perceptually relevant landmarks (A3)
- prosodic convergence
  - pitch accent and boundary types, distributions
  - pitch accent and boundary shapes (PaIntE parameters) (A1)
  - dynamic adaptation of speech rate (z-scores of segment durations) (Schweitzer/Möbius 2003)