Speaking in the presence of background speech

Martin Cooke, Vincent Aubanel, Julian Villegas, María Luisa García Lecumberri
Purpose

• Task-based dialogues somewhat artificial
  – Limited range of turn-taking initiatives
  – Real conversations may be much richer in strategies

• Conversations provide a broader perspective of how ‘noise’ affects everyday speech beyond word-level intelligibility
  – e.g. ability to detect and respond to turn-yielding cues
‘Coffee bar’ scenario

5 minutes, pair 1 present

10 minutes, both pairs present

5 minutes, pair 2 present

- Sennheiser ME-3 head-mounted microphones
- [also: AKG 4500 table mic.]
- Visual barrier between but not within pairs
- 3 sessions
- Participants given list of topics (‘holidays’, ‘shopping’,...) but allowed to talk about anything
- Most participants knew each other
Annotation

Tiers
- **Speech/nonspeech**: incl. speechlike (laughter, purposeful inbreath, etc.)
- **Words**: pseudo-phonemic
- **Events**: elongations, dysfluencies, mistimings, self-cutoff, alignment
- **Overlaps**: pauses, latches, inter-turn gaps
- **Turn-types**: smooth change, back-channel, interruption, competitive start, butting in

Approx. 5000 annotations per 20 minute session
Overlaps: within dialogue pair

- Dialogue ‘precision’ measured via cross-correlation of speech activity in pair of channels: lag at 0 shows actual overlap

- Some consistency in pairs across sessions (colours indicate same pairs in different sessions)

- Wide variety in overlap tolerated by different pairs (5-25%), and some differences in sharpness of dialogue tuning

- No clear effect of background pair present/absent
Overlaps across pairs

No evidence of reduction of overlap at zero lag - these do not look like dialogues
Why no overlap reduction in these conversations?

• Recall reduction in overlap with background from 41% to 29% in Sudoku task in presence of competing speech

• Some differences:
  – visual information not available in Sudoku task
  – task: problem-solving allows ‘thinking’ pauses while free conversation doesn’t?
  – masker level higher in Sudoku task (headphone delivery, no head-shadow)
  – Masking speech material came from other Sudoku dialogues, hence more informational masking potential
‘Standard’ Lombard effects I

As function of number of active talkers, relative to one active alone

Norm: energy compensation for other talkers

Energy, F0, F1 all indicate small Lombard effect

But speech rate reduction is much larger than predicted on basis of other parameters

<table>
<thead>
<tr>
<th>Noise level during production (dB SPL)</th>
<th>dB increase in speech production level</th>
<th>% duration increase</th>
<th>% increase in spectral C of G</th>
<th>% increase in semitone of F0</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>7.1</td>
<td>3.3</td>
<td>32</td>
<td>1.7</td>
</tr>
<tr>
<td>89</td>
<td>8.3</td>
<td>5.3</td>
<td>34</td>
<td>2.0</td>
</tr>
<tr>
<td>96</td>
<td>9.5</td>
<td>7.6</td>
<td>38</td>
<td>2.5</td>
</tr>
</tbody>
</table>
‘Lombard-like’ effects within dialogue

- Increase in energy & F0 when pair engaged in dialogue re. arbitrary pair
- Very substantial reduction in speech rate within dialogue when both active
- Suggests that Lombard effects operate even within dialogues due to interlocutor overlap
- OR: could be ‘holding the floor’ strategies (need to split into old vs new speaker)
Event counts

Increase when background present

Change in events/minute when background speaker pair present

DYS = dysfluency
MIS = mistiming
CUT = self cutoff
ELO = elongation
ALI = alignment

- Clear increase in dysfluencies
- Increase in mistimings for speakers physically closest to other pair (‘middle pairs’)
Turn type counts

Gravano & Hirschberg (2010) scheme

SS = competitive start
T = smooth turn change
BC = back-channel
I = interruption
PI = interruption during pause
BI = “butting in”

- Increase in interruptions
- Decrease in back-channel usage
Latching: rapid ‘precision’ turn exchanges
Some overlap is normal
Inter-turn ‘gaps’

• negative gaps are overlaps between turns
• in each case, increase in width of distribution
Classes of inter-turn gap

- Evidence for reduced latching ability when background pair present => less precision timing
- Increase in long overlaps and long pauses when background pair present => reduced ability to respond to turn-yielding cues (also divided attention)
Discussion

Main findings
• Much larger reduction in speech rate than would be predicted on basis of normal Lombard effect
• Strong time-domain Lombard effects operate even within dialogues due to interlocutor overlap
• Background speaker pair lead to increase in dysfluencies and possible increase in mistimings, longer overlaps; also increase in interruptions
• Evidence for reduced precision timing of turn changes
• Tendency for longer inter-turn pauses: background monitoring or missed turn-yielding cues?

BUT
• As yet no clear evidence that speakers are making things easier for listeners by retiming their contributions
Conclusions

• Natural conversations provide a real-world testbed for speech perception in noise, going beyond word intelligibility to other socially-important elements such as recognition of turn-yielding cues

• But natural conversations exhibit different constraints to task-oriented dialogues beloved of speech researchers:
  – no task imperative, hence no overt information transfer pressure
  – social constraints e.g. no extensive thinking time allowed

Planned for early Feb. in Edinburgh
  1. Replication of subset of earlier (Sudoku) study
  2. Comparative & larger collection of conversational speech with/without background dialogues in English.