The Importance of Prosody for TRP Projection

Wieneke Wesseling, Rob van Son

Institute of Phonetic Sciences
ACLC
University of Amsterdam

Nederlandse Vereniging voor Fonetische Wetenschappen
Summer Seminar on Prosody
Motivation

Question:

What is the relative importance of different prosodic features in language processing?
Motivation

**Question:**

What is the relative importance of different prosodic features in language processing?

**Method:**

*Reaction Time experiment:* Identify end-of-utterance in different conditions
Task

Minimal Response Task:

Identification of TRP’s in Dialogue
Task

Minimal Response Task:

Identification of TRP’s in Dialogue

- TRP: Transition Relevance Place
  (a new speaker can start)
- identify when to start speaking
- by saying 'AH'
- 'elicited minimal responses'
Materials

Stimulus Set
Materials

Stimulus Set

- informal dialogs from the Spoken Dutch Corpus (CGN)
- 7 telephone & 11 face-to-face dialogues (165 min.)
- basic annotation, word by word phonetic transcription
- hand aligned word boundaries (utterance endings)
- End intonation: high/mid/low
- Pitch accents: last prominent word
- Manipulated versions to test importance of various prosodic measures
Recording Setup
Response recording from laryngograph and microphone

1. 2 minute Practise set

stimulus (mono) 'Aaa'
stimulus (mono)

stimulus (stereo)
Recording Setup
Response recording from laryngograph and microphone

1. 2 minute Practise set
2. 10 x 6 minute Stimulus-files
Recording Setup

Response recording from laryngograph and microphone

1. 2 minute Practise set
2. 10 × 6 minute Stimulus-files
3. randomized for presentation
Recordings
Example response waveform and segmentation

Top: Mono waveform of the stimulus file
Tier 1: laryngograph signal
Tier 2: Automatic response segmentation, RT=yellow interval
Bottom: Annotation tiers for the utterances
The importance of prosody for TRP projection

Introduction

Experiment

Materials

Response Collection

RT model

Stimuli

Results

Discussion

Conclusions

Reaction-Times

Perception-Central-Motor model

\[
\text{Perception } t_p \quad \rightarrow \quad \text{Central} \quad \sigma \sqrt{\tau^3} \quad \alpha \quad \tau = \frac{1}{\alpha} \quad \rightarrow \quad \text{Motor } t_m
\]
Perceptual and Motor component, deterministic response-time

\[ t_0 = t_p + t_m \]
Central **decision making component** determined by an integration-time }\tau = \frac{1}{\alpha}{
Stimuli

Four sets of stimulus files:
## Stimuli

Four sets of stimulus files:

1. Original

<table>
<thead>
<tr>
<th>Stimulation Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>No prosody added</td>
</tr>
<tr>
<td>Hummed</td>
<td>Only intonation and pause structure</td>
</tr>
<tr>
<td>Whispered</td>
<td>No periodic sounds, but audible prosody</td>
</tr>
<tr>
<td>Masked</td>
<td>One or none of the last four words replaced by white noise</td>
</tr>
</tbody>
</table>

Three Experiments:

- **Experiment 1**: Original vs. Hummed (21 subjects)
- **Experiment 2**: Original vs. Whispered (11 subjects)
- **Experiment 3**: Masked vs. Non Masked (24 subjects)
### Stimuli

**Four sets of stimulus files:**

1. Original  
2. Hummed: nothing but intonation and pause structure
Stimuli

Four sets of stimulus files:

1. Original
2. Hummed: nothing but intonation and pause structure
3. Whispered: no periodic sounds (but audible prosody)
### Stimuli

Four sets of stimulus files:

1. Original
2. Hummed: nothing but intonation and pause structure
3. Whispered: no periodic sounds (but audible prosody)
4. Masked: either one or none of last four words replaced by white noise
Stimuli

Four sets of stimulus files:

1. Original
2. Hummed: nothing but intonation and pause structure
3. Whispered: no periodic sounds (but audible prosody)
4. Masked: either one or none of last four words replaced by white noise

Three Experiments:
Stimuli

Four sets of stimulus files:

1. Original
2. Hummed: nothing but intonation and pause structure
3. Whispered: no periodic sounds (but audible prosody)
4. Masked: either one or none of last four words replaced by white noise

Three Experiments:

Experiment 1 Original vs Hummed (21 subjects)
### Stimuli

**Four sets of stimulus files:**

1. Original
2. Hummed: nothing but intonation and pause structure
3. Whispered: no periodic sounds (but audible prosody)
4. Masked: either one or none of last four words replaced by white noise

### Three Experiments:

**Experiment 1** Original vs Hummed (21 subjects)

**Experiment 2** Original vs Whispered (11 subjects)
Stimuli

Four sets of stimulus files:

1. Original
2. Hummed: nothing but intonation and pause structure
3. Whispered: no periodic sounds (but audible prosody)
4. Masked: either one or none of last four words replaced by white noise

Three Experiments:

Experiment 1  Original vs Hummed (21 subjects)
Experiment 2  Original vs Whispered (11 subjects)
Experiment 3  Masked vs Non Masked (24 subjects)
Subjects can predict an upcoming TRP from *high* or *low* boundary tones

but, most likely, have to wait until they perceive the end of the utterance (pause) in *mid* boundary tone *intonation only* stimuli
Results Experiment 2
Original vs Hummed & Whispered Speech

- Clear correlation between average RT and distance to last accent in all but hummed stimuli. Responses to hummed utterances are only affected by the final accent.
- Accent position affected the response delays of intelligible speech.
Results Experiment 3: RT Distribution

Distribution of RTs with respect to Position of Masked word.

Distribution of elicited responses comparable to natural turn-switches.
Results: Response distribution

<table>
<thead>
<tr>
<th>Position</th>
<th>Presentations</th>
<th>Responses</th>
<th>Prob. of Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>last</td>
<td>13,066</td>
<td>4,194</td>
<td>0.321</td>
</tr>
<tr>
<td>last-1</td>
<td>9,538</td>
<td>3,594</td>
<td>0.377</td>
</tr>
<tr>
<td>last-2</td>
<td>8,558</td>
<td>3,358</td>
<td>0.392</td>
</tr>
<tr>
<td>last-3</td>
<td>7,650</td>
<td>2,984</td>
<td>0.390</td>
</tr>
<tr>
<td>No Mask</td>
<td>25,627</td>
<td>7,216</td>
<td>0.282</td>
</tr>
<tr>
<td>Total</td>
<td>64,439</td>
<td>21,346</td>
<td>0.331</td>
</tr>
</tbody>
</table>

- 24 hrs/64,439 utt., 24 subj. eliciting 21,436 responses
- Only utterances adding 'content'. (no minimal responses, grounding acts, 'formulaic' expressions, interjections, hesitations, coughs and unintelligible speech)
- All statistics done after correcting for subject identity
Results: Mask Position vs Utterance Length

Mean Reaction-time delays for different Mask pos. wrt. utterance length.

- Effect for utterance length for all utterances
- Specifically for Unmasked utterances vs. Masked pooled
- Individual Mask positions did not differ significantly
Results: Relative Delays 1
Relative RT delays for Length and Mask position.

- **relative RT** for Mask pos.& length, compared to unmasked utterances of same length
- effect of masking for last word of utterance, not for masked words before the last word.
Results: Relative Delays 2
Relative RT delays for Length and Mask position.

possible exception for single word utterances,(unresolved)
Results: Standard Deviations

SD for different Mask positions with respect to utterance length.

- Utter. with length 1 or 2, larger SD for masked
- For longer utterances, same trend but not significant.
- No effect for length on SDs
Results: Prominence, Relative RTs

Mean relative RT for position of last prominent word and Mask position

- No effect of the position of the prominent word
- Only when penultimate word is prominent and ultimate word masked, responses slower
Results: Prominence, Standard Deviations

SD of relative RT for position of last prominent word and Mask position (utterance length > 2).

- All masked utterances pooled have longer SD (and thus longer integration times)
- This effect is found for all mask positions (after, on or before the prominent word)
Discussion: Masking

- Number & delays of our subjects’ responses primarily determined by utterance length
- Only if the last word was masked, masking had effect on RT
- Masking did have an effect on processing efforts
Discussion: Prominence

- Whether or not the last prominent word was masked had no effect on reaction times.
- Only when penultimate word was prominent & last word masked, delay in RT was found.
Conclusions

- The presence of -even prominent- masked words, did not affect RTs unless the last word of the utterance was masked.
Conclusions

- The presence of -even prominent- masked words, did not affect RTs unless the last word of the utterance was masked.
- Only to determine the exact position of the TRP, is the identity of the last word needed.
The presence of -even prominent- masked words, did not affect RTs unless the last word of the utterance was masked.

Only to determine the exact position of the TRP, is the identity of the last word needed

Identity of the last word before a TRP is used to predict the timing of the response
Conclusions

- The presence of even prominent masked words, did not affect RTs unless the last word of the utterance was masked.
- Only to determine the exact position of the TRP, is the identity of the last word needed.
- Identity of the last word before a TRP is used to predict the timing of the response.
- Predicting the relative position of the last word before the TRP is robust enough to be unaffected by missing individual words.
Conclusions

- The presence of -even prominent- masked words, did not affect RTs unless the last word of the utterance was masked.
- Only to determine the exact position of the TRP, is the identity of the last word needed.
- Identity of the last word before a TRP is used to predict the timing of the response.
- Predicting the relative position of the last word before the TRP is robust enough to be unaffected by missing individual words.
- The strong facilitating effect of utterance length on RTs also points to the use of global syntactic and discourse structure in predicting the relative position of the last word.
Response Probability
Probability of response by Mask position & Utterance length

<table>
<thead>
<tr>
<th>Position of Mask</th>
<th>Length of utterance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>last</td>
<td>0.195</td>
<td>0.262</td>
</tr>
<tr>
<td>last-1</td>
<td>-</td>
<td>0.280</td>
</tr>
<tr>
<td>last-2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>last-3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No Mask</td>
<td>0.225</td>
<td>0.270</td>
</tr>
</tbody>
</table>

- Subjects more likely to respond to unmasked utterances
- Subjects more likely to respond to longer utterances