ABSTRACT
Speech is considered an efficient communication channel. This implies that the organization of utterances is such that more speaking effort is directed towards important parts than towards redundant parts. Based on a model of incremental word recognition, the importance of a segment is defined as its contribution to word-disambiguation. This importance is measured as the \textit{segmental information content}, in bits. On a labeled Dutch speech corpus it is then shown that crucial aspects of the information structure of utterances partition the segmental information content and explain 90\% of the variance. Two measures of acoustical reduction, duration and spectral center of gravity, are correlated with the segmental information content in such a way that more important phonemes are less reduced. It is concluded that the organization of conventional information structure does indeed increase efficiency.

THE IMPORTANCE OF A PHONETIC SEGMENT

- **Lexical Information Content** $I_L$ (bits)
  - Phonemic contribution to word recognition based on an incremental word recognition model

- **Segmental Information Content** $I_S$ (bits)
  - $I_S$ corrected for average word predictability in context based on \textit{Context Distinctiveness}

FORMULA'S

**Lexical Information Content** $I_L$

$$I_L = - \log \left( \frac{\text{Frequency} \left( \text{word onset} \mid s \right)}{\text{Frequency} \left( \text{word onset} \mid \text{any segment} \right)} \right)$$ (based on incremental word recognition)

**Context Distinctiveness of a word w: $\text{CD}(w)$**

$$\text{CD}(w) = \sum_{\text{vowels}} p(c \mid w) \log \frac{p(c \mid w)}{p(c)}$$

- $\text{KL}(P(c) \mid P(c|w))$ word lag
- $P(c\mid w)$: Probability of context word c in the neighborhood of w
- $P(c)$: Probability of c in general

**Segmental Information Content** $I_S$

(i.e. average in context)

Define: $D(w) = \text{RelFreq}(w) \cdot (2^{\text{CD}(w)} - 1)$

$$I_S = - \log \left( \frac{\text{Frequency} \left( \text{word onset} \mid s \right) + D(w)}{\text{Frequency} \left( \text{word onset} \mid \text{any segment} \right) + D(w)} \right)$$

EXAMPLE: /o/ in Dutch 'boom' (English 'tree')

- Relative CGN frequency of boom: 5.05 \times 10^{-3}
- Context Distinctiveness: $\text{CD}(\text{boom}) = 4.53$
- Relative frequency in context: $2^{\text{CD}(\text{boom})-5.05\times10^{-3}} = 1.2 \times 10^{-3}$
- Original smoothed CELEX word count of boom: 2,226 occurrences
- Context-corrected CELEX count: 45,402 (1.2 \times 10^{-3} \times 39 - 10^6)
- Correction term (eq. 3): 45,402 - 2,226 = 43,176
- Words starting with /bo:/: 67,710 (1.172 CELEX entries)
- Words starting with /b/\.: 1,544,483 (26,186 CELEX entries)

\[ I_L = \log \left( \frac{67,710}{154,4483} \right) = 4.51 \text{ (eq. 1)} \]
\[ I_S = \log \left( \frac{45,402 \cdot 1,544,483}{67,710 \cdot 43,176} \right) = 3.84 \text{ (eq. 4)} \]

\( \Rightarrow I_L < I_S \) context reduces lexical uncertainty.

INTRODUCTION

- **Prosodic and Phonetic Features of Utterances Reflect Information Structure (i.e. Importance)**
- **Speech is Efficient**:
  - Important Entities are Emphasized
  - Redundant Entities are De-emphasized

Examples:
- New Concepts are put in Focus and at the End
- Function Words are Redundant, Short, Reduced, and Never in Focus

INFORMATION STRUCTURE AND EFFICIENCY

CENTRAL QUESTION:

How are Redundancy and Reduction distributed at the Segmental Level?

AIMS:

- Quantify the Importance of Linguistic Factors to the Distribution of Information at the Phoneme Level
- Link Information Structure and Phonetic Reduction

MATERIALS & METHODS

- **CELEX Word-Frequency list**
  - 38 Million Words

- **Spoken Dutch Corpus**
  - 1.8 Million Words, 5th rel.

- **IFA corpus**
  - 8 speakers, 50,000 Words

Explained Variance:

Maximal Reduction of "Within Factor" Variance after Adding the Factor

Acoustic Measures of Reduction:

- **Duration**
- **Spectral Center of Gravity**
  - First Spectral Moment (\textit{all} phonemes)
  - **Formant Contrast**

Distance between a vowel/realization in $F_1$ and $F_2$, formant space (in semitones) and a virtual target of reduction (each speaker separately). Reduction of a vowel results in a shorter distance to this virtual point in vowel space.
Linguistic Factors Explored

<table>
<thead>
<tr>
<th>Segment</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phoneme position</td>
</tr>
<tr>
<td>2</td>
<td>Phoneme identity</td>
</tr>
<tr>
<td>3</td>
<td>Nr. of Syllables</td>
</tr>
<tr>
<td>4</td>
<td>Prominence</td>
</tr>
<tr>
<td>5</td>
<td>Lexical Stress</td>
</tr>
<tr>
<td>6</td>
<td>Cluster length</td>
</tr>
<tr>
<td>7</td>
<td>Syllable Part</td>
</tr>
<tr>
<td>8</td>
<td>Word position</td>
</tr>
<tr>
<td>9</td>
<td>Syllable position</td>
</tr>
</tbody>
</table>

Contributions to Vowel $I_L$ & $I_S$ with respect to segmental factors (100%)

Contributions to Vowel Duration and Formant Contrast with respect to segmental factors (100%)

Contributions to Variance of Duration

Contributions to Variance of the Spectral Center of Gravity

Conclusions

- Information Structure is measurable down to the Segmental Level
- Acoustic Reduction is Aligned with Information Structure
- Variation is Distributed in an Efficient Way

But:

- There is a lot of "noise", meaning that we have missed important factors
- Larger (and better) corpora are needed