Information Structure and Efficiency in Speech Production

R.J.J.H. van Son and Louis C.W. Pols
Institute of Phonetic Sciences/ACLC
University of Amsterdam, the Netherlands
{Rob.van.Son, Louis.Pols}@hum.uva.nl

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 *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_L corrected for average word predictability in context based on Context Distinctiveness

FORMULA'S

Lexical Information Content I_L

$$I_L = -\log_2\left(\frac{Frequency([word\ onset]+s)}{Frequency([word\ onset]+any\ segment)}\right)$$

(based on incremental word recognition)

Context Distinctiveness of a word w: *CD(w)*

$$CD(w) = \sum_{vocabulary} P(c_i|w)\log_2\frac{P(c_i|w)}{P(c_i)}$$

Kullback-Leibler distance between $P(c_i)$ and $P(c_i|w)$ (use [-5,5] word-bag)
$P(c_i|w)$: Probabillity of context word c_i in the neighbourhood of w
$P(c_i)$: Probabillity of c_i in general

Segmental Information Content I_S (i.e. average in context)

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

$$I_S = \log_2\left(\frac{Frequency([word\ onset]+s)+D(w)}{Frequency([word\ onset]+any\ segment)+D(w)}\right)$$

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

Relative CGN frequency of *boom*:	$5.05\cdot10^{-5}$
Context Distinctiveness:	*CD(boom)* = 4.53
Relative frequency in context:	$2^{CD(boom)}\cdot5.05\cdot10^{-5} = 1.2\cdot10^{-3}$
Original smoothed CELEX word count of *boom*:	2,226 occurrences
Context-corrected CELEX count:	45,402 ($1.2\cdot10^{-3}\cdot39\cdot10^{6}$)
Correction term (eq. 3):	*D(boom)*= 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$_2$(67710/1544483) = **4.51** (eq. 1)
➢ I_S=-log$_2$([67710+43176]/[1544483+43176]) = **3.84** (eq. 4)
==> $I_S < I_L$ 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

FORM FOLLOWS ***FUNCTION***

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

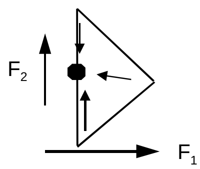
MATERIALS & METHODS

- **CELEX** Word-Frequency list (38 Million Words)
  I_L Lexical Information
- **Spoken Dutch Corpus** (1.8 Million Words, 5th rel.)
  I_L==> I_S Segmental Information
- **IFA corpus** (8 speakers, 50,000 Words)
  Labeled & Segmented Speech, Segments, Reduction
  Informal and Read speech

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

Acoustic Measures of Reduction:

- Duration
- Spectral Center of Gravity
  First Spectral Moment (*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.

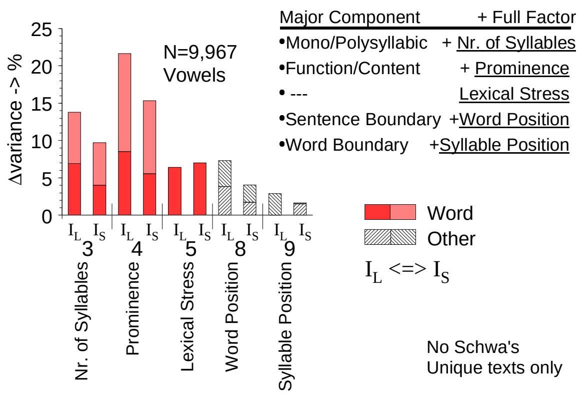


LINGUISTIC FACTORS EXPLORED

SEGMENT	1. Phoneme position	: Position of Segment in Word
	2. Phoneme	: Phoneme Identity
WORD	**3. Nr. of Syllables**	: Word-length in Syllables
	4. Prominence	: Automatic Prominence (0-4)
	5. Lexical Stress	: Lexical Syllable Stress
SYLLABLE *Consonants*	6. Cluster length	: Length of Consonant Clusters
	7. Syllable Part	: Onset, Kernel, or Coda
OTHER	8. Word position	: Position of Word in Sentence
	9. Syllable position	: Position of Syllable in Word

CONTRIBUTIONS TO *VOWEL* I_L&I_S

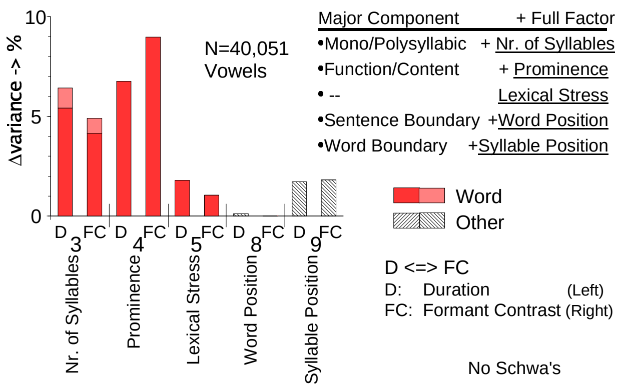
WITH RESPECT TO SEGMENTAL FACTORS (100%)



Variance after Segmental Factors (1&2) = 100% *Note*: Linear Scale

CONTRIBUTIONS TO *VOWEL* **D**URATION AND **F**ORMANT **C**ONTRAST

WITH RESPECT TO SEGMENTAL FACTORS (100%)



Variance after Segmental Factors (1&2) = 100% *Note*: Linear Scale

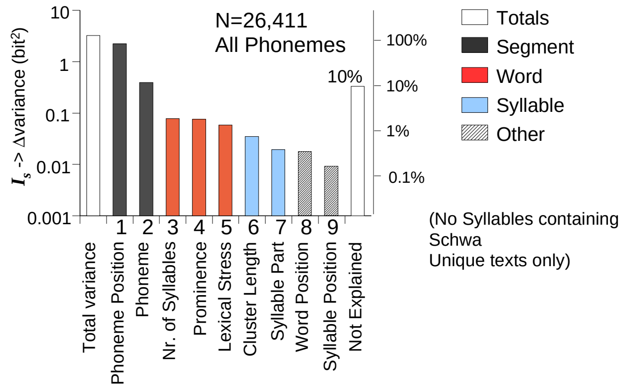
CONCLUSIONS

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

BUT:

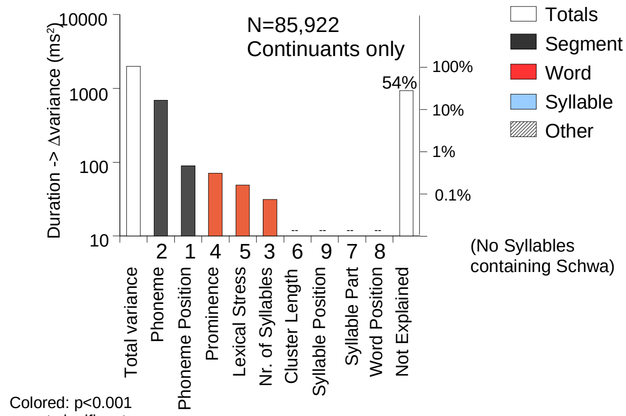
- There is a lot of "noise", meaning that we have missed important factors
- Larger (and better) corpora are needed

CONTRIBUTIONS TO THE VARIANCE OF SEGMENTAL INFORMATION I_S



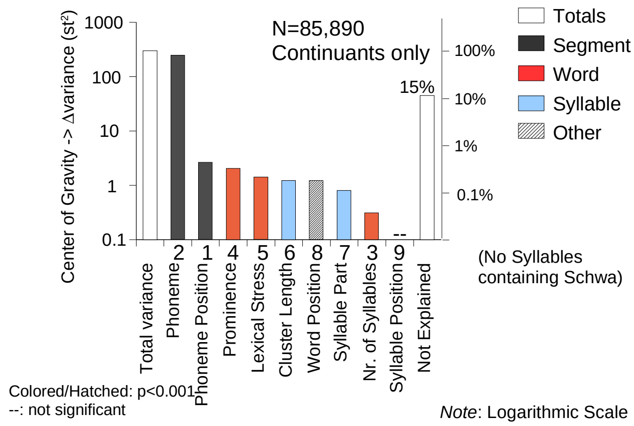
Note: Logarithmic Scale

CONTRIBUTIONS TO VARIANCE OF DURATION



Note: Logarithmic Scale

CONTRIBUTIONS TO VARIANCE OF THE SPECTRAL CENTER OF GRAVITY



Note: Logarithmic Scale