Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010).
Signal-based and expectation-based factors in the perception of prosodic prominence.
Laboratory Phonology, *1*(2), 425-452.

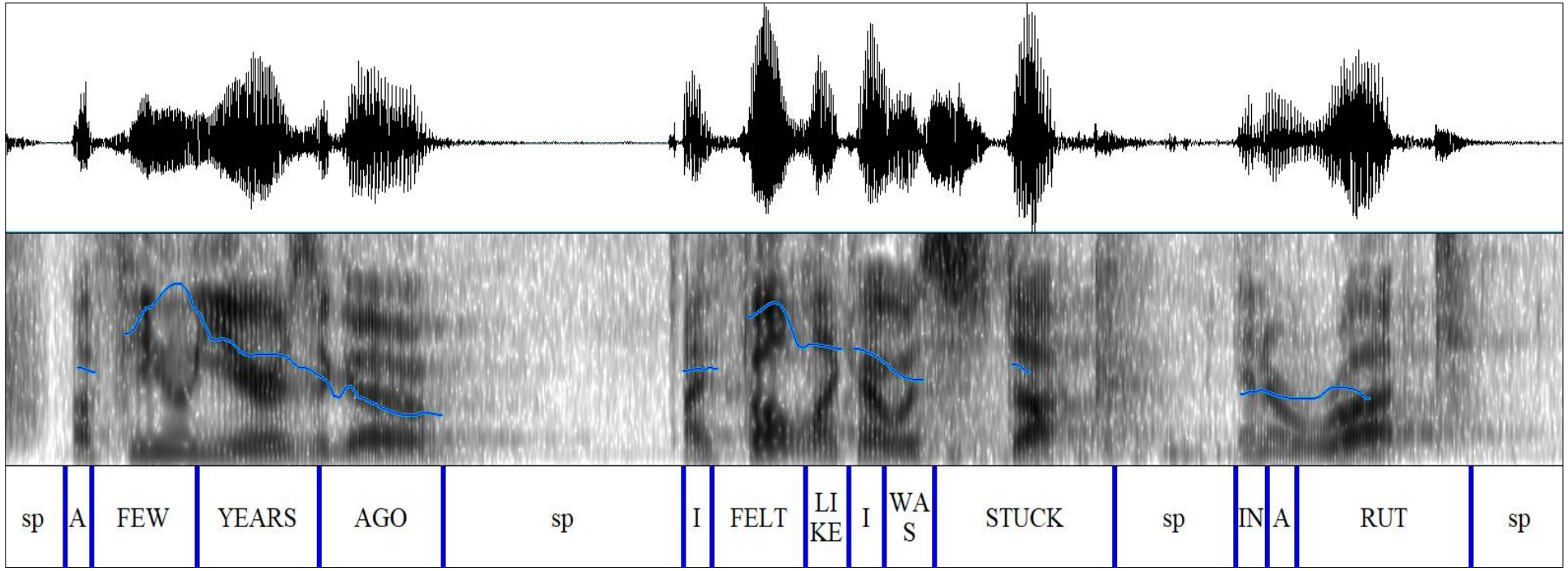
한국음운론학회 강독회

2020. 10. 17. (토)

임수연 (한양대학교, suyeonim03@gmail.com)

1. Introduction

- The spoken form of a word
  - Vowels and consonants
  - Prosodic context

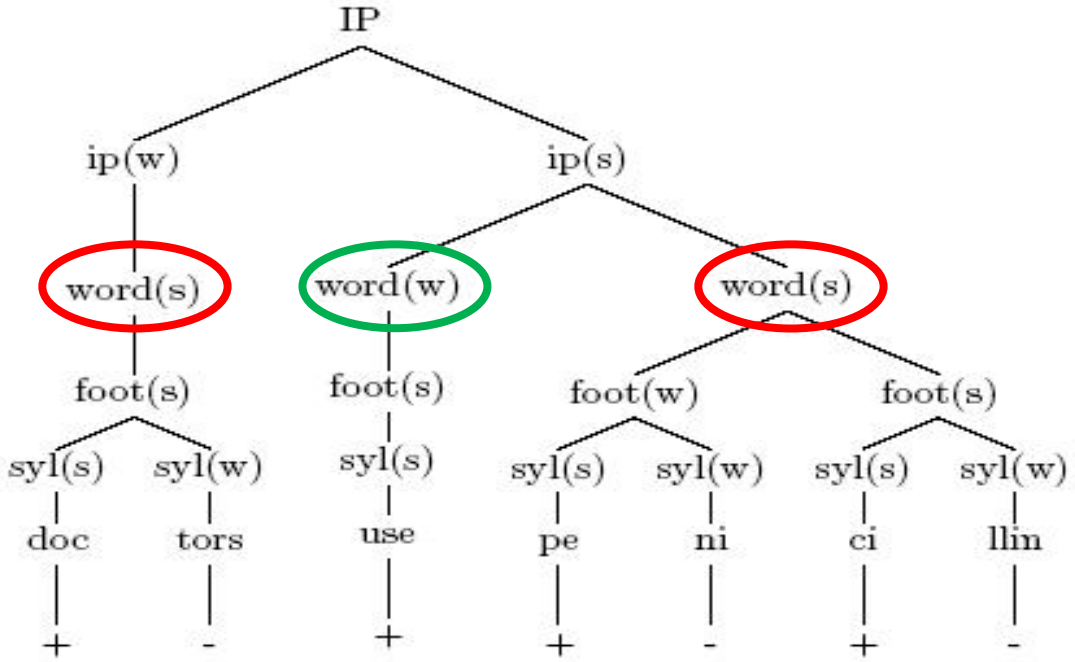


1. Introduction

- In this paper,
  - Production and perception of prosodic context in ordinary speech
  - Specifically, prosodic prominence (the strength of a spoken word relative to the words surrounding it in the utterance)

1. Introduction

- A syllable or word is prominent if…
- It is parsed in a strong position in metrical structure (“structural” prominence)



https://en.wikipedia.org/wiki/Metrical_phonology

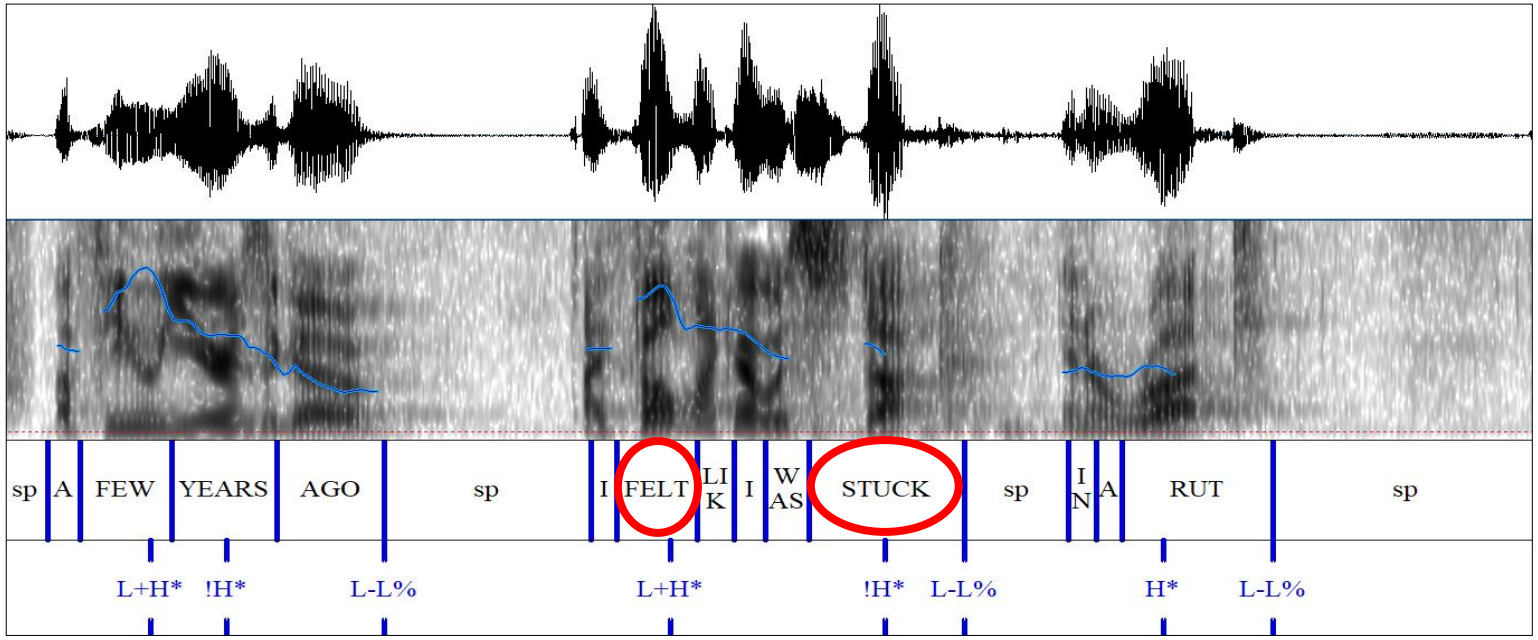
1. Introduction

- A syllable or word is prominent if…
- It introduces new or important information in discourse.
  - e.g. President Obama delivered a brilliant speech in Tucson.
- Or, it bears contrastive focus.
  - e.g. Did you call John? No, I called Mary.

Baumann & Riester (2013); Baumann et al. (2006)

1. Introduction

- Prominence is reflected in the phonetics (“acoustic” prominence).
  - hyper-articulation
  - increased duration and intensity
  - a salient F0 movement, etc.

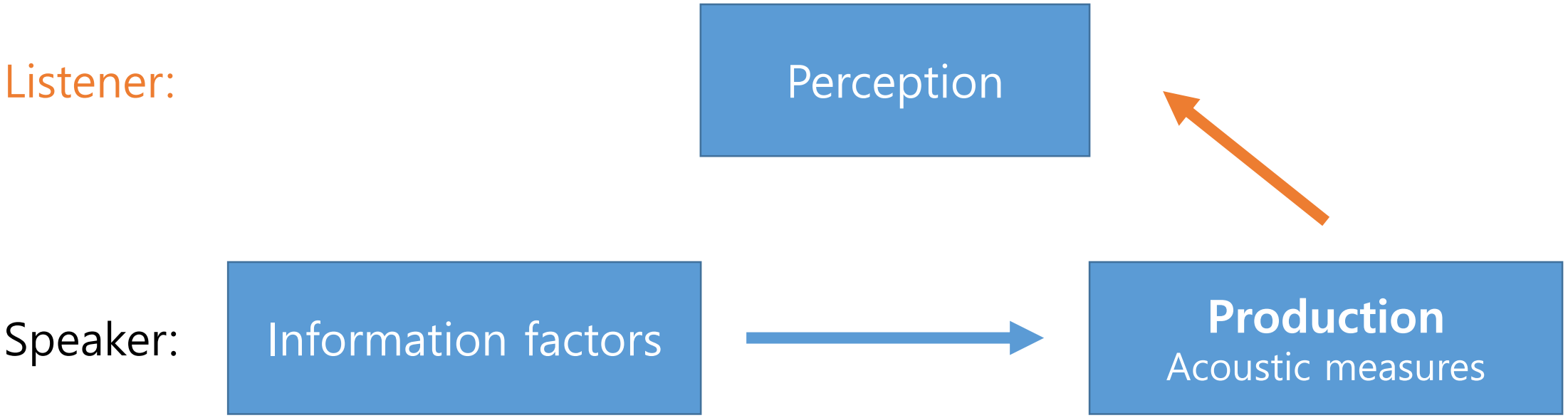


1. Introduction

- **Speakers** assign structural prominence, taking into account discourse properties of the words, that is realized as acoustic prominence through increased duration, intensity, and F0 patterns.

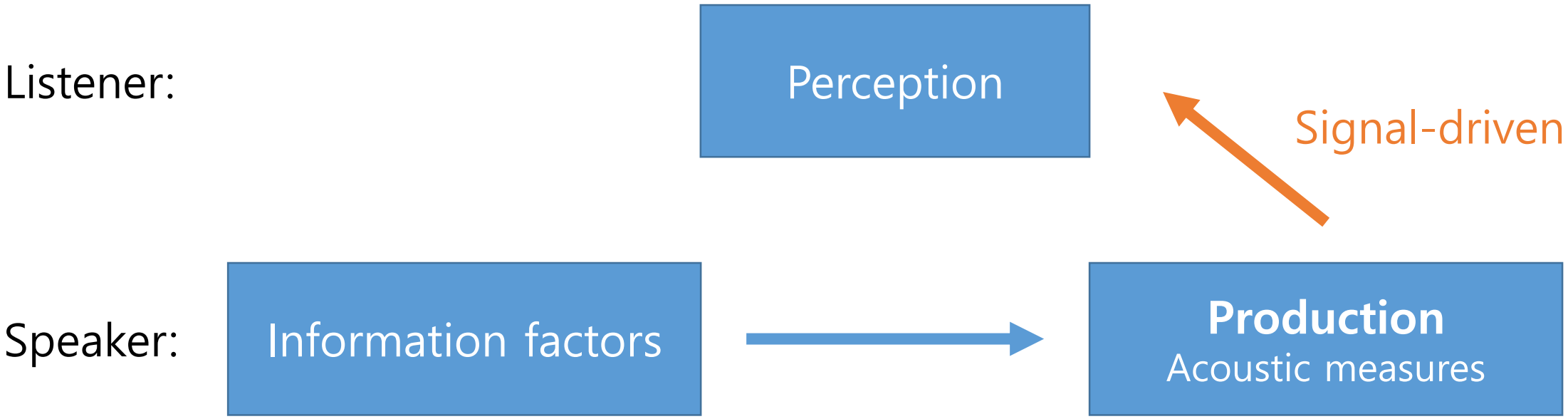
1. Introduction

- **Listeners** perceive words with acoustic prominence as referring to new entities in discourse or entities with contrastive focus, while words with less acoustic prominence in association with prior discourse context.



1. Introduction

- Simple model of **signal-driven** prosody perception
- However, **additional considerations** that complicate the model

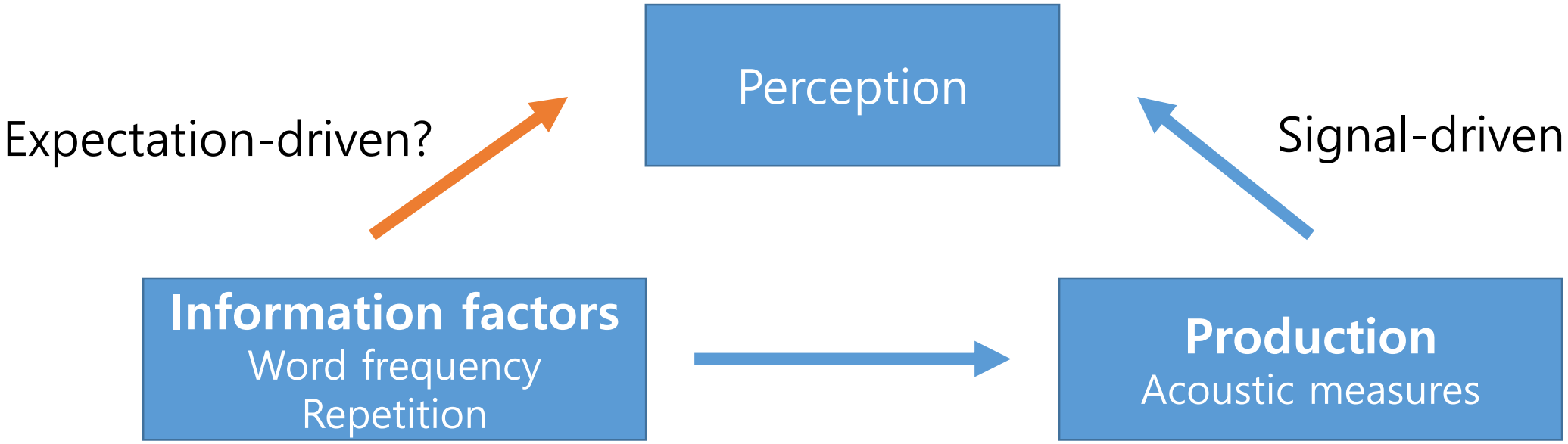


1. Introduction

- Words that are **predictable** from the discourse context have reduced acoustic prominence.
- High **frequency** words exhibit a greater incidence of consonant lenition and vowel reduction than low-frequency words.

1. Introduction

- Bard and Aylett (1999): Words that are repeated have reduced acoustic prominence as expected, but are often **still perceived as structurally prominent**.

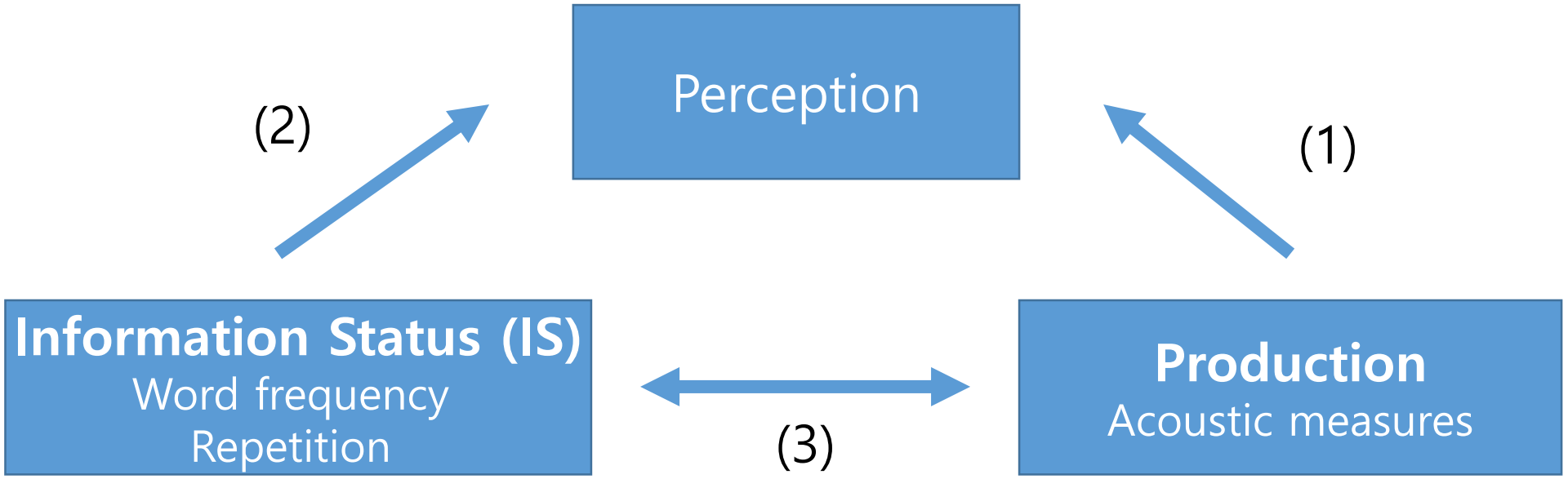


1. Introduction

- **How do listeners perceive prominence in everyday speech?**
- Ordinary and spontaneous speech
  - Broad range of expressions and focus conditions
  - Richness in the phonetic reduction
- Listeners' perception of prominence broadly construed
- These will be discussed in the next Section 2.

1. Introduction

- Prominence in relation to:
  - (1) Acoustic correlates (Section 3.1)
  - (2) Words' information status (IS) (Section 3.2)
  - (3) The relationship between acoustic correlates and IS (Section 4)



1. Introduction

- Section 5:
  - More extended statistical analysis
  - **IS, word repetition and frequency,** influence prominence perception at least partly **independently of the acoustic properties** of a word.
- Section 6:
  - A processing model
  - Prominence perception is **both signal-driven**, based on the speaker's phonetic implementation of prominence, and **expectation-driven**, based on the listener's prior experience.

2. Experiment in naïve prosody transcription

- To explore the correlates of prominence in acoustic or information status, it is necessary to construct **annotation for the location of prominence**.
- Often conducted by a small number of highly **trained experts**
- This may differ from ordinary, **non-linguistic expert listeners**.
- Coarse prosody annotation by **97 untrained transcribers of American English** from University of Illinois

2.1. Materials

- Buckeye corpus of conversational speech (Pitt et al. 2007)
- 90 excerpts of **interviews** (37 speakers)
  - Short excerpts (11-22 seconds in duration)
  - Long excerpts (31-58 seconds)
- Transcriptions were modified to **remove all punctuation and capitalization**.

2.2. Method

- Subjects seated at computers with the instruction:

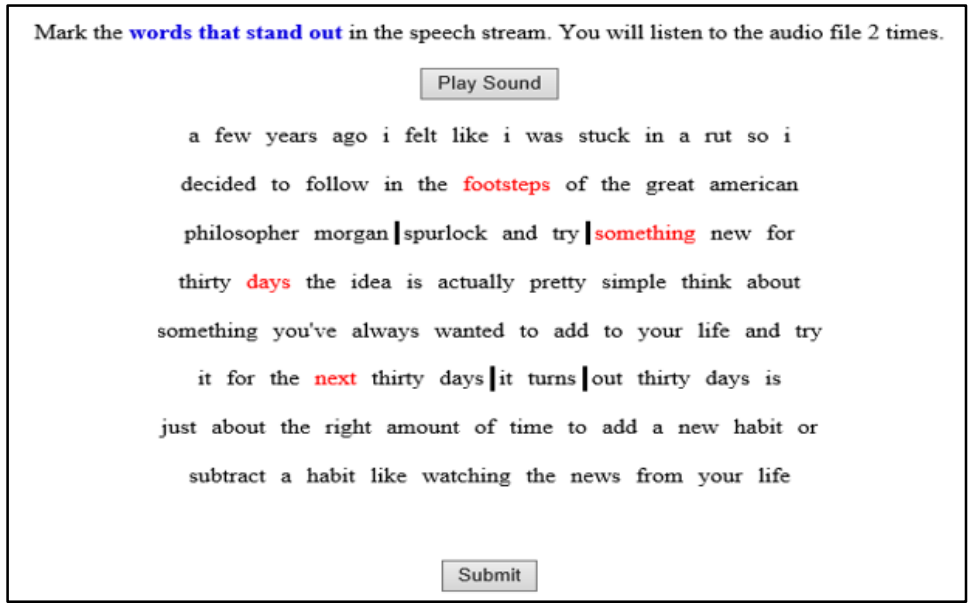
"In normal speech, speakers pronounce some word or words in a sentence with more prominence than others. **The prominent words are in a sense highlighted for the listener, and stand out from other non-prominent words.** In some of the excerpts you will hear, you will be asked to mark all prominent words by underlining them."

2.2. Method

- Subjects listened to the speech experts twice through headphones and were asked to mark the printed transcript for the location of prominent words in real time.

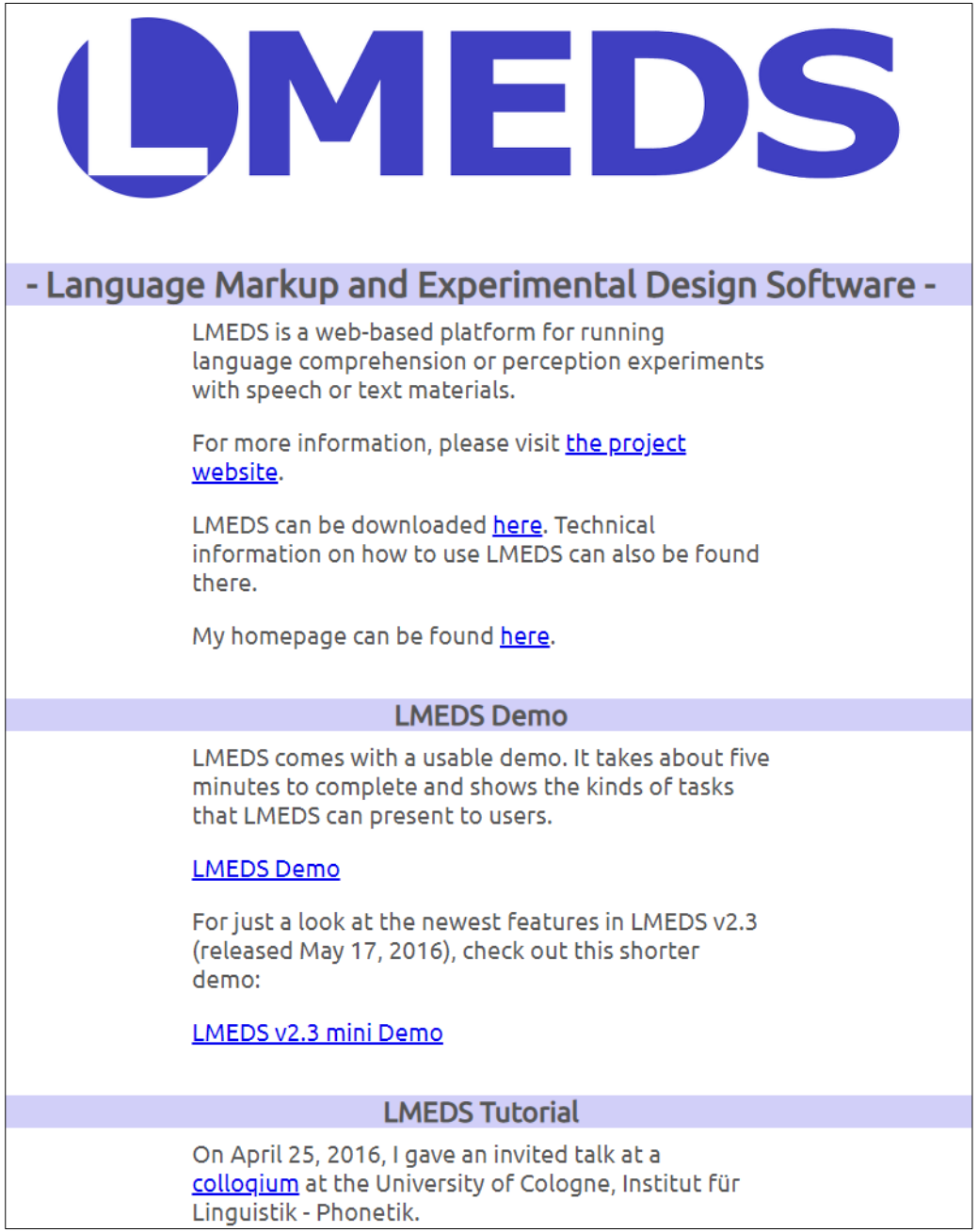
(2) a. word word word
b. word word word
c. word word word

LMEDS (Language Markup and Experimental Design Software; Mahrt, 2016)



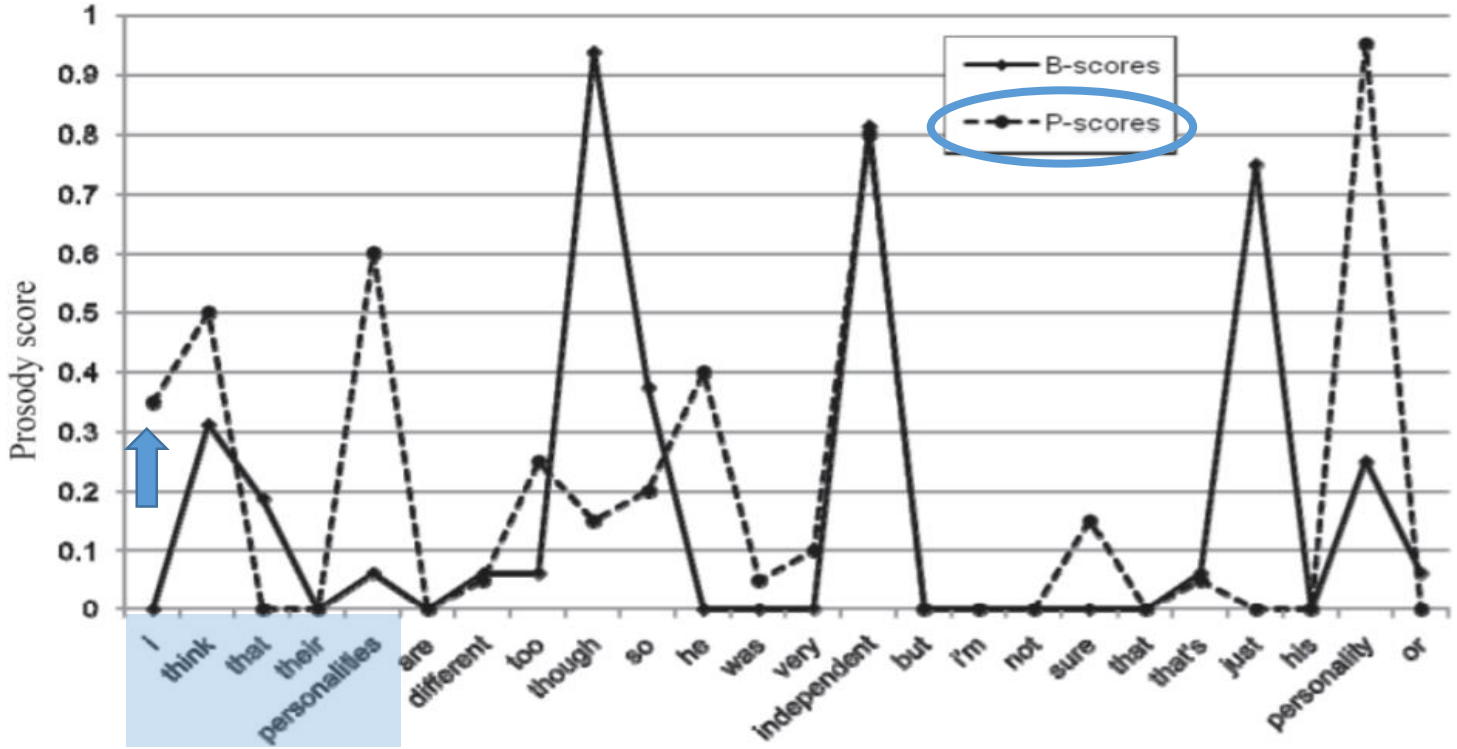
2.2. Method

- Tim Mahrt (2016) https://www.timmahrt.com/lmeds.html



2.3. Data coding and assessing reliability

- Prominence (p-) score: a number between 0 and 1 that represents the proportion of transcribers who perceive that word as prominent.



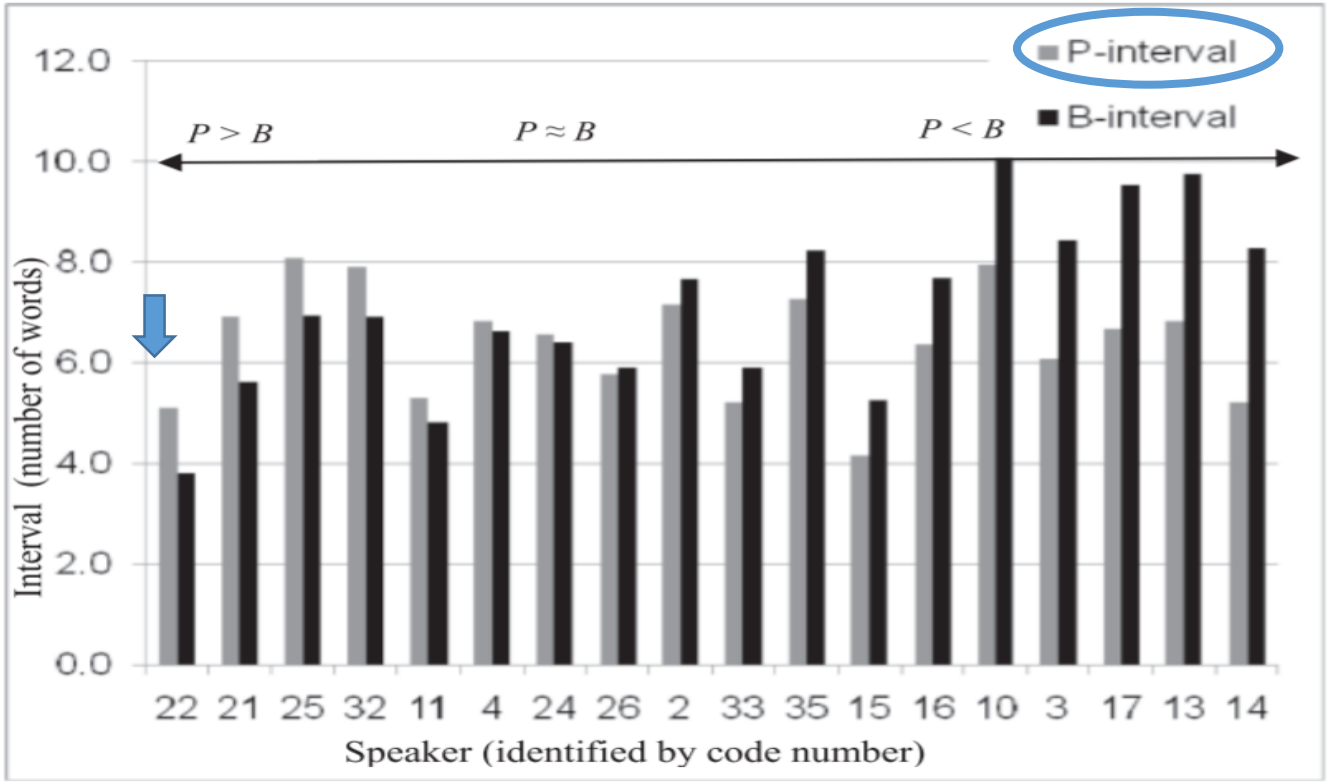
2.3. Data coding and assessing reliability

- **Variability of listeners** in their perception of prominence:
  - Fleiss' kappa coefficients
  - Their normalized *z*-scores
- All z-scores are highly significant (at $\alpha = 0.01$, $z = 2.32$)
- **Listeners agree at above chance levels.**

Excerpt set		1	2	3	4
prominence	Kappa	0.373	0.421	0.394	0.407
$\alpha = 0.01$, $z = 2.32$	z	19.43	20.48	18.15	18.31
boundary	Kappa	0.612	0.544	0.621	0.575
	z	27.62	21.87	25.05	26.22

2.3. Data coding and assessing reliability

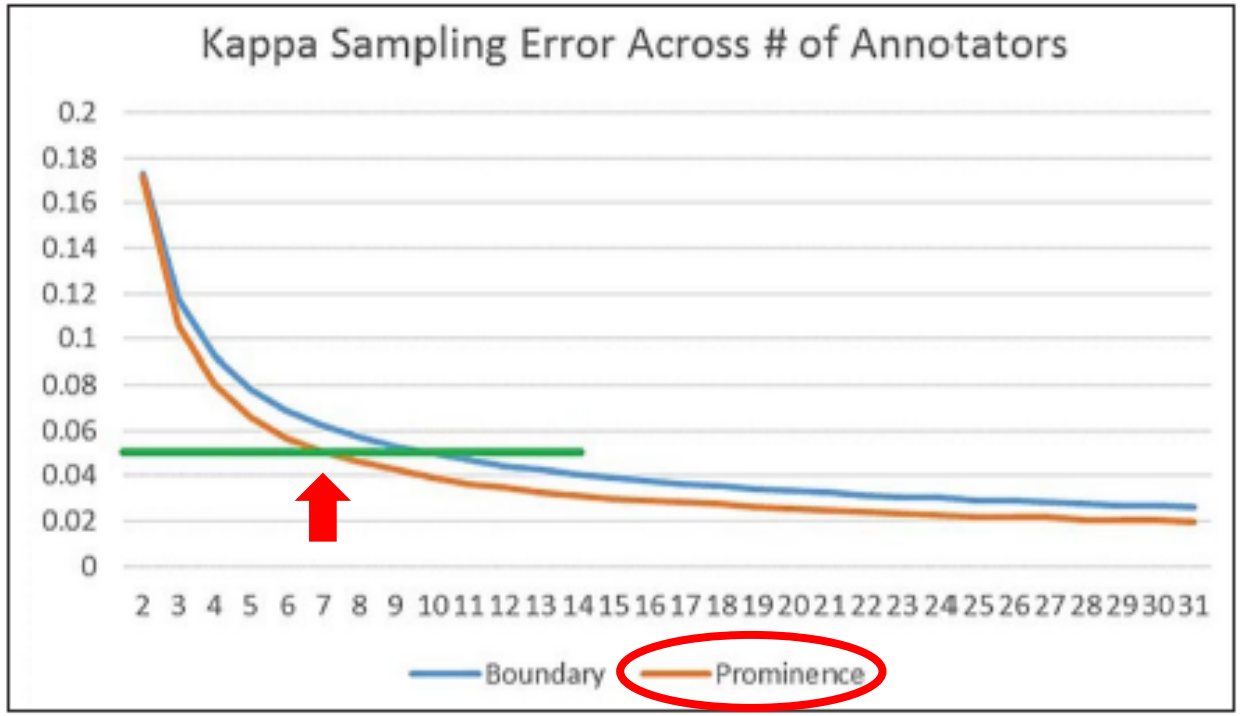
- **Variability of speakers** in their use of prosody
- P-interval: A measure of the frequency of prominent words



2.3. Data coding and assessing reliability

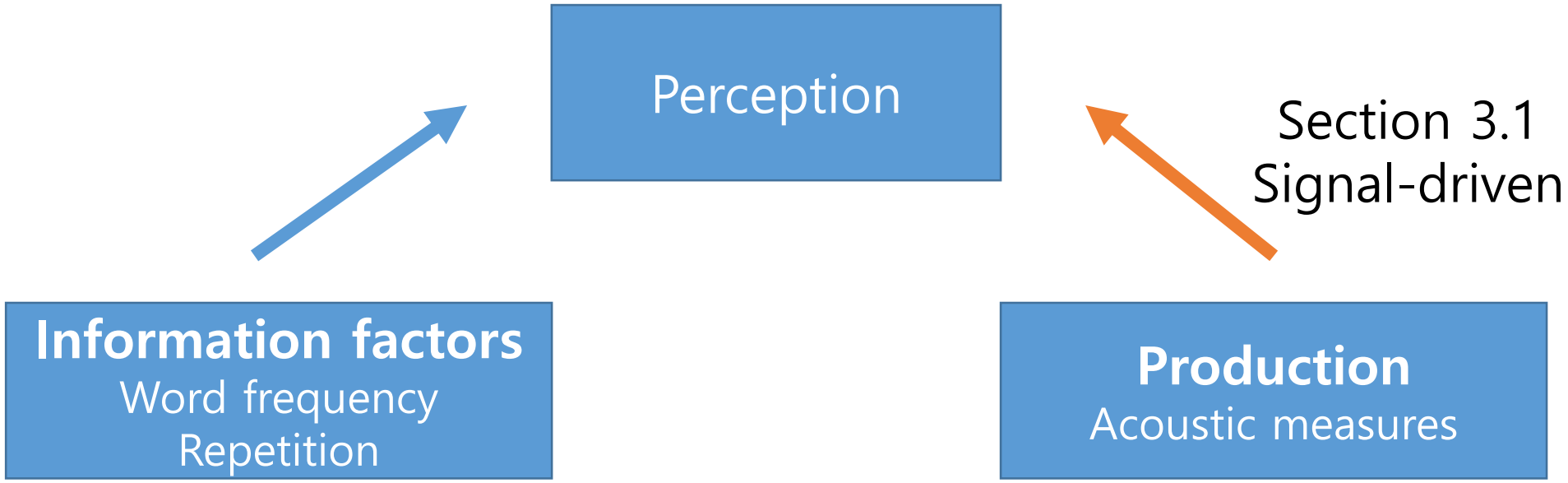
- Roy, Cole & Mahrt (2017): Using **Rapid Prosody Transcription (RPT)**, **7 annotators** yields consistent **prominence** rating (orange line)

  (c.f. Green line: an estimated sampling error of .05)



3. Correlates of perceived prominence

- **Prominence perception is signal-driven.**
- **Acoustic properties** of words in relation to their perceived prominence



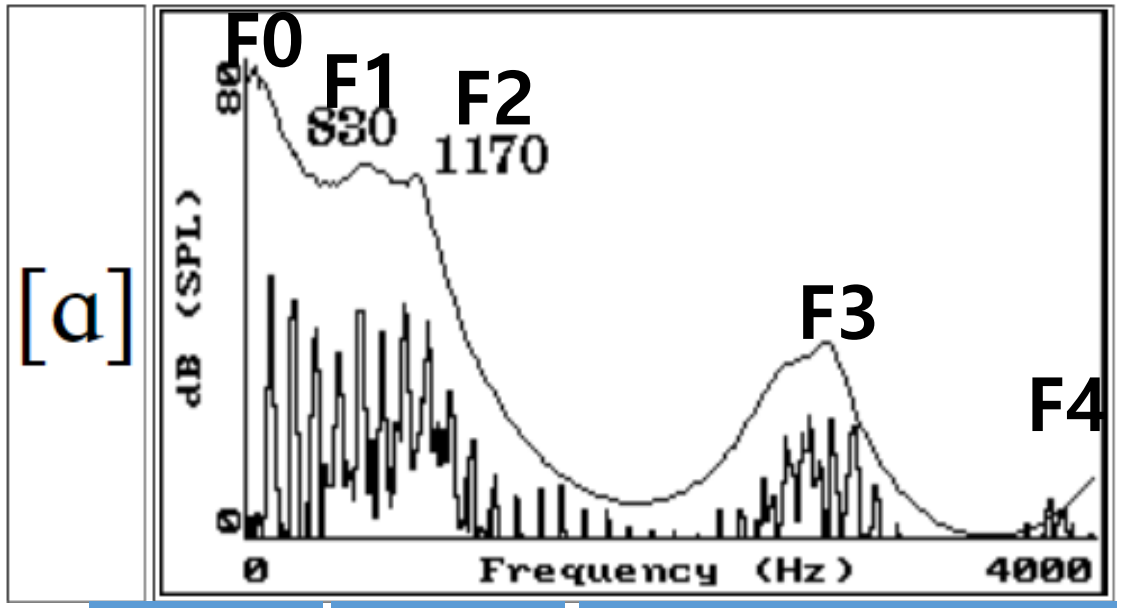
3.1. Acoustic differences related to perceive prominence

- For stressed vowels of each word,
- *Z*-normalized acoustic measures
  - **Duration** (ms)
  - **Overall** RMS **intensity** (dB)
  - **Spectral emphasis**: RMS intensity (dB) in different frequency bands → A measure of increased vocal effort
- Correlation between acoustic measures and prominence
- Consistency with previous studies

vowel	ɑ	æ	ʌ	ɔ	aʊ	aɪ	ɛ	ɝ	eɪ	ɪ	i	oʊ	ʊ	u
N	81	129	211	58	28	140	187	66	114	209	156	103	41	94

3.1. Acoustic differences related to perceive prominence

- Sluijter & van Heuven (1996): Stressed vowels are produced with more vocal effort, If a speaker produces more vocal effort, higher frequencies increase more than lower frequencies.



0-.5 kHz (low)	.5-2 kHz (mid)	2-4 kHz (high)

F4https://home.cc.umanitoba.ca/~krussll/phonetics/acoustic/formants.html

3.1. Acoustic differences related to perceive prominence

- For stressed vowels of each word,
- *Z*-normalized acoustic measures
  - **Duration** (ms)
  - Overall RMS **intensity** (dB)
  - Spectral emphasis: RMS **intensity** (dB) in different frequency bands → A measure of increased vocal effort
- Correlation between acoustic measures and prominence
- Consistency with previous studies

vowel	ɑ	æ	ʌ	ɔ	aʊ	aɪ	ɛ	ɝ	eɪ	ɪ	i	oʊ	ʊ	u
N	81	129	211	58	28	140	187	66	114	209	156	103	41	94

3.1. Acoustic differences related to perceive prominence

- Significant **positive** correlation between acoustic measures and p-scores
- Greater correlation strength for **duration than intensity**
- No correlation for non-low back, rounded vowels

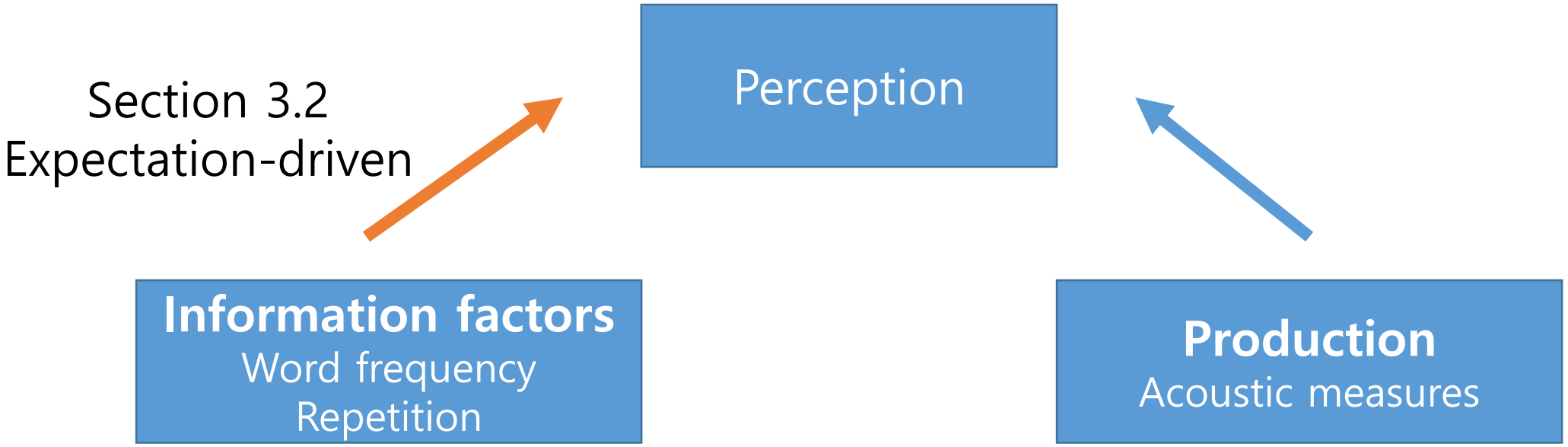
vowel	ɑ	æ	ʌ	ɔ	aʊ	aɪ	ɛ	ɝ	eɪ	ɪ	i	oʊ	ʊ	u
Duration	.033	.301**	.198**	.224*	.491*	.419**	.237**	.160	.302**	.244**	.266**	−.128	−.042	.141
Overall intensity	.304**	.114	.147*	−.043	.151	.209**	.220**	.283**	.137	.228**	.139*	.123	.308*	.005
Intensity 0–.5 kHz	.174	.098	.078	−.096	.076	.184*	.137*	.237*	.116	.210**	.138*	.140	.268*	.005
Intensity .5–2 kHz	.343**	.163*	.271**	.041	.209	.238**	.282**	.349**	.093	.262**	.105	.120	.328*	.024
Intensity 2–4 kHz	.175	.263**	.150*	.001	.098	.152*	.264**	.035	.184*	.201**	.132	−.018	.177	−.056

3.1. Acoustic differences related to perceive prominence

- In sum,
- Consistent results with those of previous studies
- Words with **weak acoustic cues** perceived as **non-prominent**
- Words with **enhanced acoustic cues**, especially **increased duration**, perceived as **prominent**

3. Correlates of perceived prominence

- **Prominence perception is expectation-driven.**
- **Information status** of words in relation to their perceived prominence



3.2. Information status correlates of perceived prominence

- For each word in excerpts,
- Two measures of information status
  - **Frequency**
  - **Repetition**
- Function words: Overall high frequency
- 80 frequently reduced words (Huddleston & Pullum, 2002): Potential different relationship with prominence

3.2. Information status correlates of perceived prominence

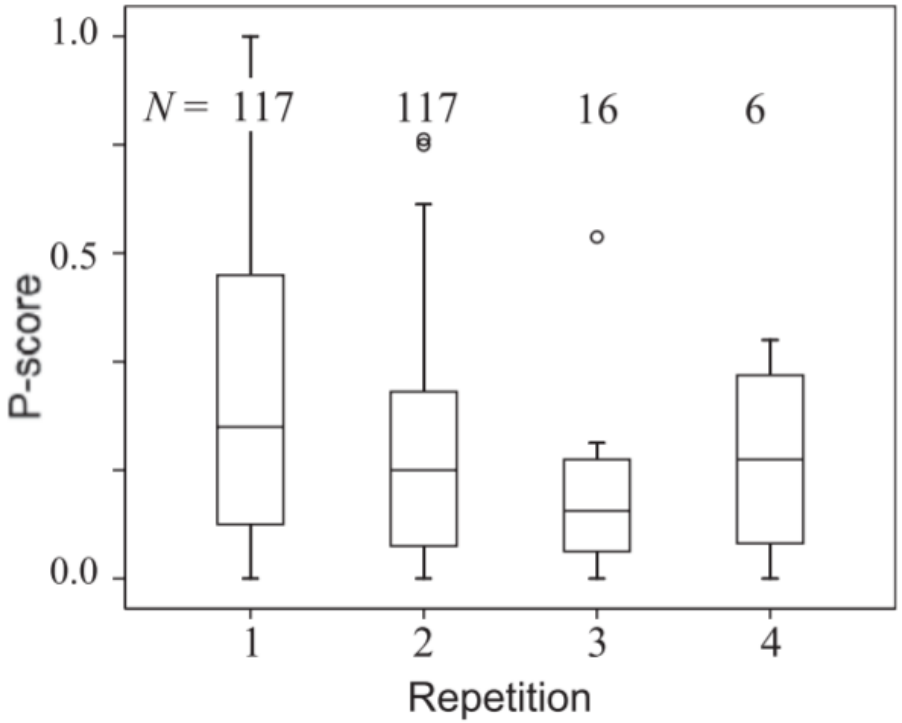
- Significant **negative** correlation between **frequency** & p-scores
- Higher predictability for all words (including function and reduced words) → Function words are reduced and perceived as non-prominent.

(1)Cont.+Func.
(2)Cont.+some func.
(3)Content only

Data set	N	Pearson's *r*	r^2
Short excerpts, all words	2024	−.505**	.255
Short excerpts minus frequently reduced words	1217	−.432**	.187
Short excerpts minus function words	778	−.302**	.091
Long excerpts	1725	−.432**	.187

3.2. Information status correlates of perceived prominence

- Relationship between **repetition** and p-scores in short excerpts
- **Decreasing** p-scores for 1-3th mention
- **Increasing** p-scores for 4th mention



3.2. Information status correlates of perceived prominence

- Significant **negative** correlation between **repetition** & p-scores
- Higher predictability for 1-3th mention than 1-6th mention → Upward trend of p-scores for 4-6th mention in the previous figure

(1)Cont.+Func.

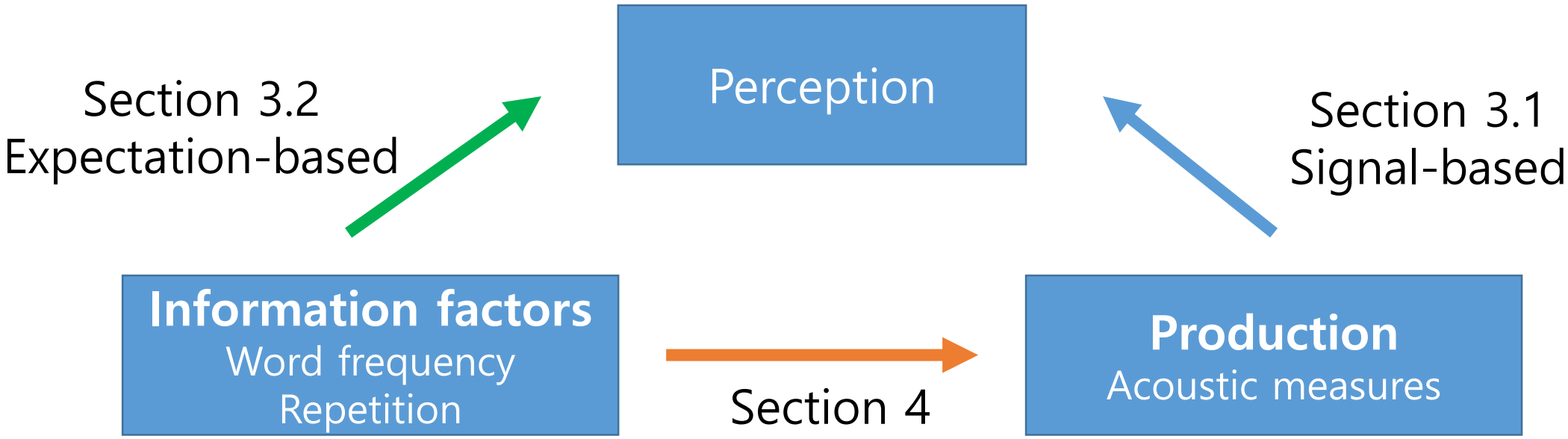
Data set	Repetition coding	N	Pearson's *r*	r^2
Short excerpts, all words	1st–6th repetition	891	−.113**	.013
	1st vs. 2nd vs. 3rd+	891	−.128**	.016
Short excerpts minus function words	1st–4th repetition	164	−.242**	.059
Long excerpts, all words	1st–5th+ repetition	481	−.061	.002
Long excerpts, all words	1st vs. 2nd repetition only	299	−.139**	.017

3.2. Information status correlates of perceived prominence

- To sum up,
- Listeners' perception of **prominence** related to word **frequency** (26%), and to a lesser extent, word **repetition** in discourse (6%)
- Function words and reduced words

3.3. Interim summary

- Does **acoustic prominence** correlate with **IS** in the corpus?



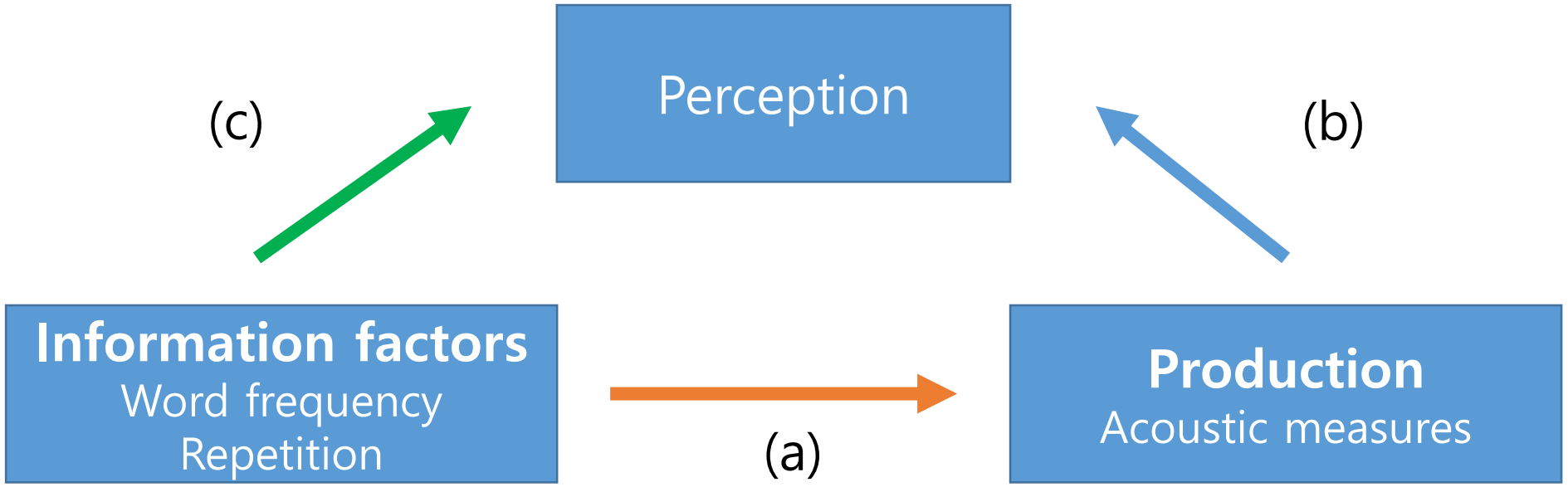
4. Acoustic correlates of word (log-) frequency and repetition

- **Complementary** relationship between **acoustic measures** and **IS**
  - Duration – repetition
  - Intensity – frequency
- **Low predictability**, except the intensity 2-4 kHz
- Word frequency > repetition

	Log-frequency	Repetition
Duration	0.001	0.024*
Overall intensity	0.024*	0.001
Intensity 0–.5 kHz	0.034*	0.002
Intensity .5–2 kHz	0.028*	0.003
Intensity 2–4 kHz	0.449*	0.002

4. Acoustic correlates of word (log-) frequency and repetition

- Are the effects of **IS** on **perceived prominence** modulated through **acoustic information**?
  - Yes: (a) = (b)
  - Probably not: (a) ≠ (b)



4. Acoustic correlates of word (log-) frequency and repetition

- Perceived **prominence** and word **frequency differ** in their **acoustic characterization**.
- The effects of word frequency on perceived prominence is **not** completely **modulated** through acoustics.
- They appear to be partly **independent** of acoustic prominence.

		Frequency	Perceived prominence
Significant acoustic correlates	duration	No	Yes ($r^2 = .089$)
	intensity	Yes ($r^2 = .024$)	Yes ($r^2 = .053$)
	spectral emphasis	Yes ($r^2 = .449$) (high frequency)	Yes ($r^2 = .07$) (mid frequency)

5. Statistical models of P-score variance

- **Individual contribution** of acoustic and IS factors in p-scores
- Hierarchical Linear Models (HLM) where factors entered in a step-wise fashion
- Models Ⅰ & Ⅱ (27%): **IS** (18-19 %) > **duration** (6 %) > **intensities** (2-3 %)

			R^2	R^2 change	Sig. of R^2 change
With acoustic measures	Model I	Log_freq & rep	.187	.187	<.001
27 %		Dur	.245	.058	<.001
		Intensities	.269	.024	<.001
	Model II	Dur	.064	.064	<.001
		Intensities	.093	.030	<.001
		Log_freq & rep	.269	.175	<.001

5. Statistical models of P-score variance

- Principle Component Analysis (PCA) to remove redundancy in the set of factors in Model Ⅰ&Ⅱ
  - PC (intensity): overall intensity & spectral emphasis
  - PC (info): frequency & repetition
- More reduced less successful (cf. Model Ⅰ & Ⅱ: 27%)

→ Individual factors' contribution

			R^2	R^2 change	Sig. of R^2 change
With PCA of intensity (26 %)	Model III	Log_freq & rep	.187	.187	<.001
		Dur	.245	.058	<.001
		PC (intensity)	.260	.015	<.001
	Model IV	Dur	.064	.064	<.001
		PC (intensity)	.080	.018	<.001
		Log_freq & rep	.260	.179	<.001
With PCA of word info (21 %)	Model V	PC (info)	.126	.126	<.001
		Dur	.185	.059	<.001
		intensities	.214	.029	<.001
	Model VI	Dur	.064	.064	<.001
		intensities	.093	.030	<.001
		PC (info)	.214	.121	<.001
With PCA of intensity and word info (20 %)	Model VII	PC (info)	.126	.126	<.001
		Dur	.185	.059	<.001
		PC (intensity)	.202	.017	<.001
	Model VIII	Dur	.064	.064	<.001
		PC (intensity)	.081	.018	<.001
		PC (info)	.202	.121	<.001

5. Statistical models of P-score variance

- The Hierarchical Linear Models use linear regression.
- **Non-linear** regression analyses to test the possibility of the non-linear relationship between factors and p-scores
- **Linear** model is more comparable and simpler than cubic model.

		R^2	Sig.
Word_freq & rep	Linear	.187	<.001
	Quadratic	.187	<.001
	Cubic	.195	<.001
Dur	Linear	.061	<.001
	Quadratic	.066	<.001
	Cubic	.070	<.001
PC (intensity)	Linear	.021	<.001
	Quadratic	.021	<.001
	Cubic	.022	<.001

5. Statistical models of P-score variance

- In sum,
- Under all the models, **IS > acoustic factors**
- All the factors are **significant** and **non-redundant**.
- Prominence perception is both **signal-driven** (influenced by acoustic factors) and **expectation-driven** (influenced by word frequency and repetition).

6. A processing model of factors influencing prominence perception

- What is the **mechanism** by which word **frequency** and **other IS** factors can influence a listener's judgment of **prominence**?
- Word predictability is inversely related to acoustic prominence.
- **Speech processing model**: Predictable words are strongly activated due to local priming or frequency in the language.
- **In production**, the processing for the **predictable** words starts sooner, giving rise to **reduced** word forms. Words that are **less predictable** lack this facilitation, and so may exhibit the **less reduced** form of the word.

6. A processing model of factors influencing prominence perception

- **Listener's perception** of prominence may directly reflect the demands of speech processing.
- When lexical access is facilitated through **high activation** levels, there are **fewer demands** on the processing resources used in speech understanding.
- A listeners may judge a word as **prominent** when processing is **resource-intensive** (e.g. low-frequency, unfamiliar, or unpredictable words).

6. A processing model of factors influencing prominence perception

- **Prominence** is both **speaker-based** and **listener-based** phenomenon.
  - Speaker-based phenomenon: Acoustic prominence can arise through lexical access in production.
  - Listener-based phenomenon: Perceived prominence can arise through the processing demands of comprehending speech.
- Very often, these two sources **converge** but this model allows when a speaker perceives a word as prominent, reflecting resource-intensive processing, even when the speaker has not produced the word with strong acoustic cues.
- Prominence perception is **signal-driven** and **expectation-driven**.

7. Conclusion

- Untrained listeners reliably perceive prosodic prominence in spontaneous speech based on their impressions from listening.
- Prominence perception is **signal-driven** (correlated with acoustic measures, especially duration) and **expectation-driven** correlated with frequency and repetition in discourse).
- The relationship between perceived prominence and word frequency is **not wholly mediated** through acoustic signal.
- **Information** measures > **acoustic** measures

7. Conclusion

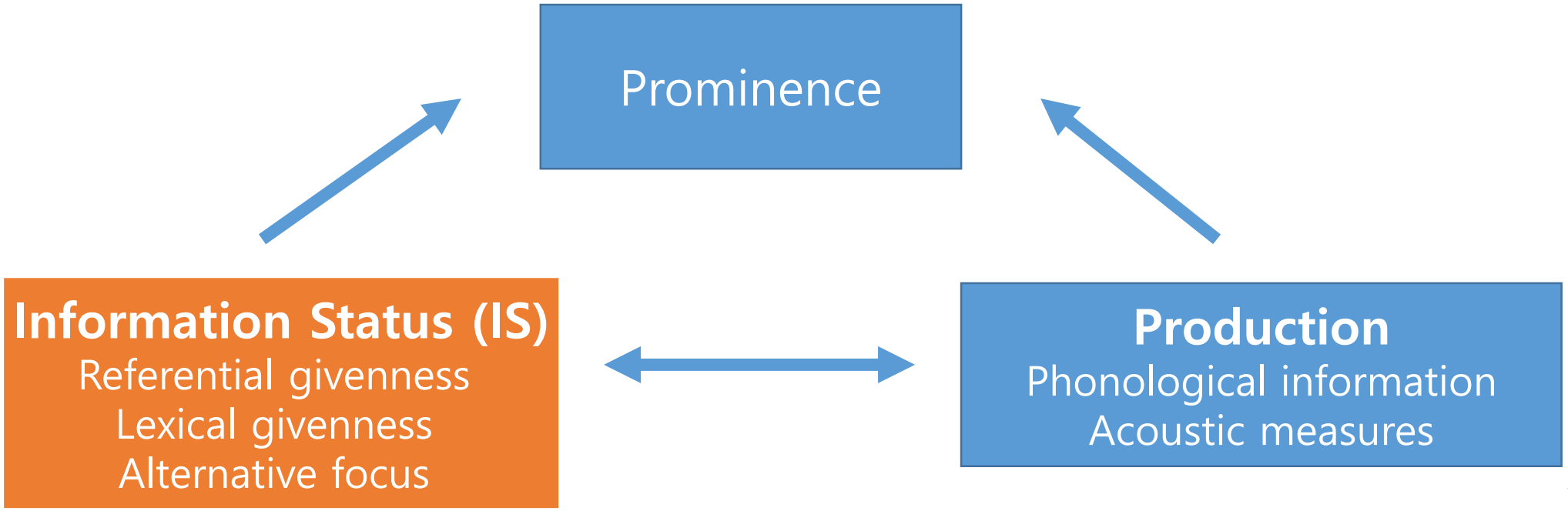
- In the processing model proposed here, **prominence perception reflects the relative attention** the listener commits to processing each word in its given discourse context.

(1) A word with acoustic prominence attracts listener's attention in direct response to the acoustic modulation.

(2) A relatively unpredicted word demands greater attention because of the lower activation levels of its (sub-) lexical units.

More recent study (Im, Cole, & Baumann, in review)

- Public speech style
- Word (lexical) repetition → Lexical/Referential givenness + focus



Methods: IS Annotation

- A simplified RefLex scheme (Riester & Baumann, 2017)

Smith was very optimistic.
The polls showed a solid majority for the politician.
referentially (r-) given
lexically (l-) new

- Alternative-based contrastive focus (Rooth, 2012)

*The examples adopted from Baumann & Riester (2013)

Methods: Labels for Referential (r-) Givenness

- r-new:

  A car was waiting in front of the hotel. I could see a woman in the car.

- r-unused:

  President Barack Obama delivered a brilliant speech in Tucson.

- r-bridging:

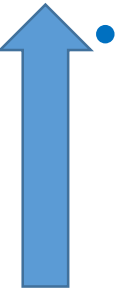
  I tried to open the door but the lock was rusty.

- r-given:

  A car was waiting in front of the hotel. I could see a woman in the car.

*The examples adopted from Baumann & Riester (2013)

Methods: Labels for Lexical (l-) Givenness & Alternative (alt) Focus



- l-new:

  A car was waiting in front of the hotel. I could see a woman in the car.

- l-given:

  A car was waiting in front of the hotel. I could see a woman in the car.

- alt (semantic alternative):

  Did you call John? No, I called Mary.

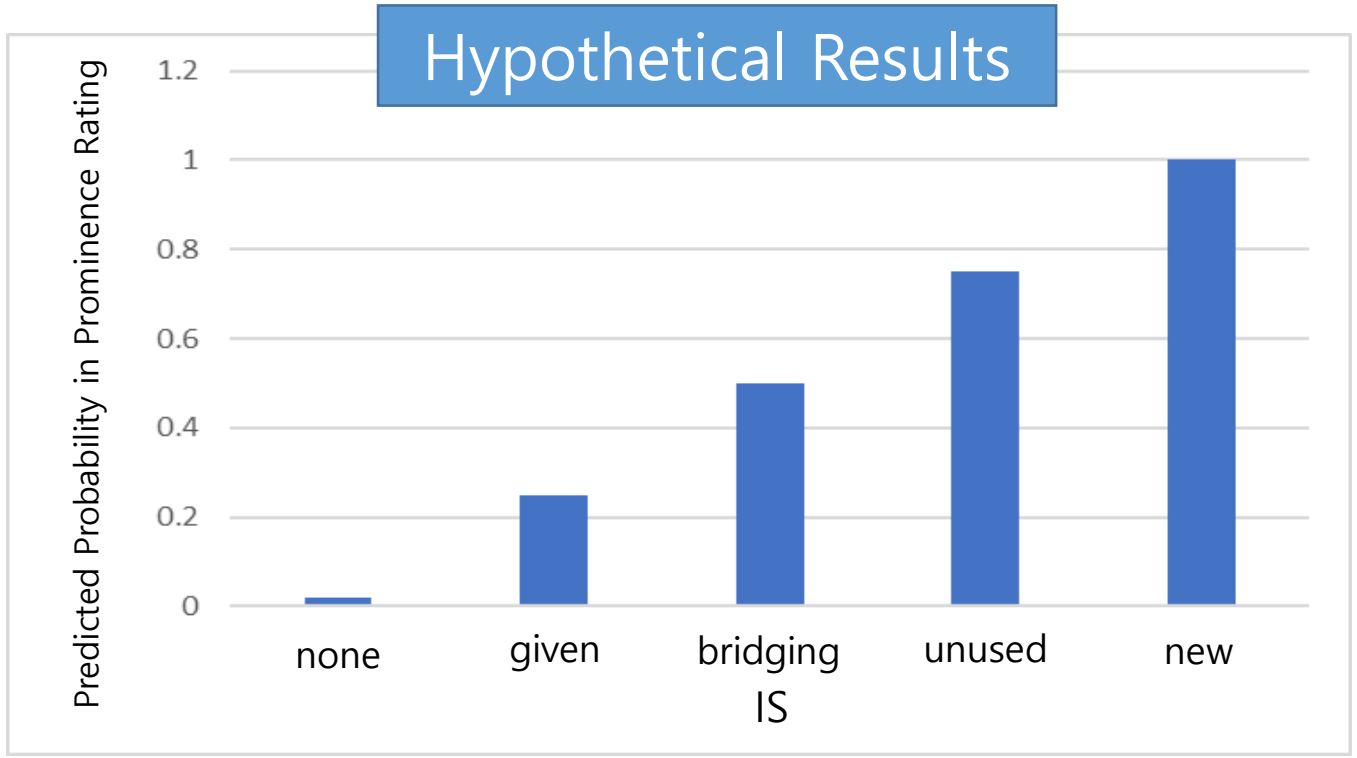
*The examples adopted from Baumann & Riester (2013); Baumann et al. (2006)

Methods

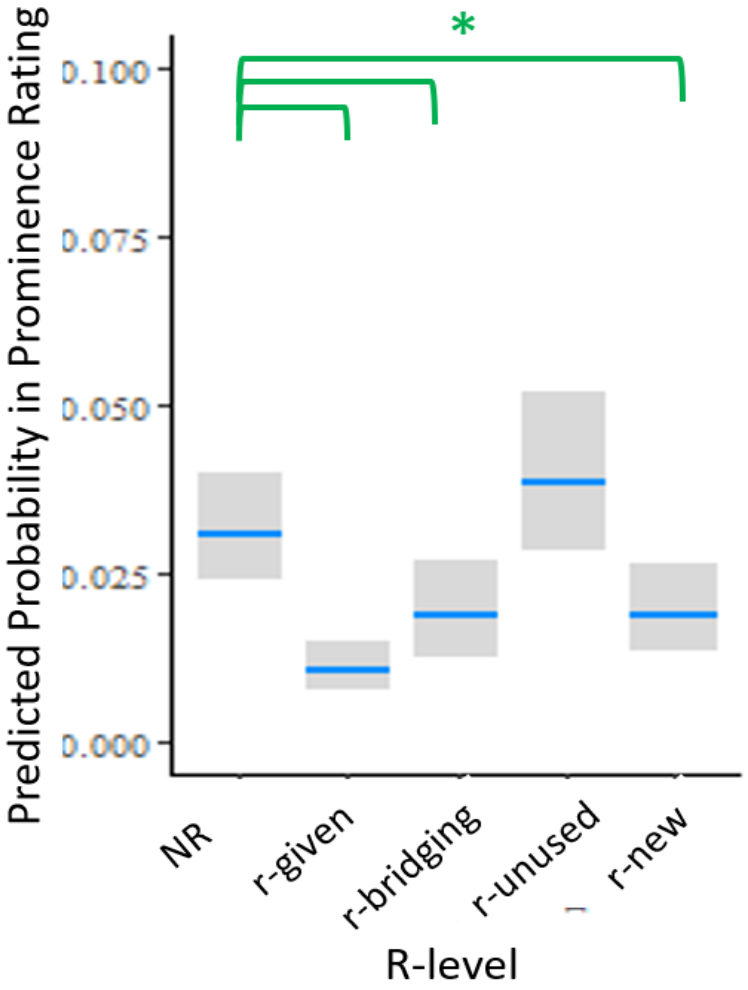
- Rapid Prosody Transcription (Cole et al., 2010)
- 35 listeners of American English
- A public speech was obtained from TED Talk delivered by a male speaker of American English in a clear and engaging manner

Prediction

- The words with new information or contrastive meaning are more likely to be perceived as prominent than words with given information or non-contrastive meaning.



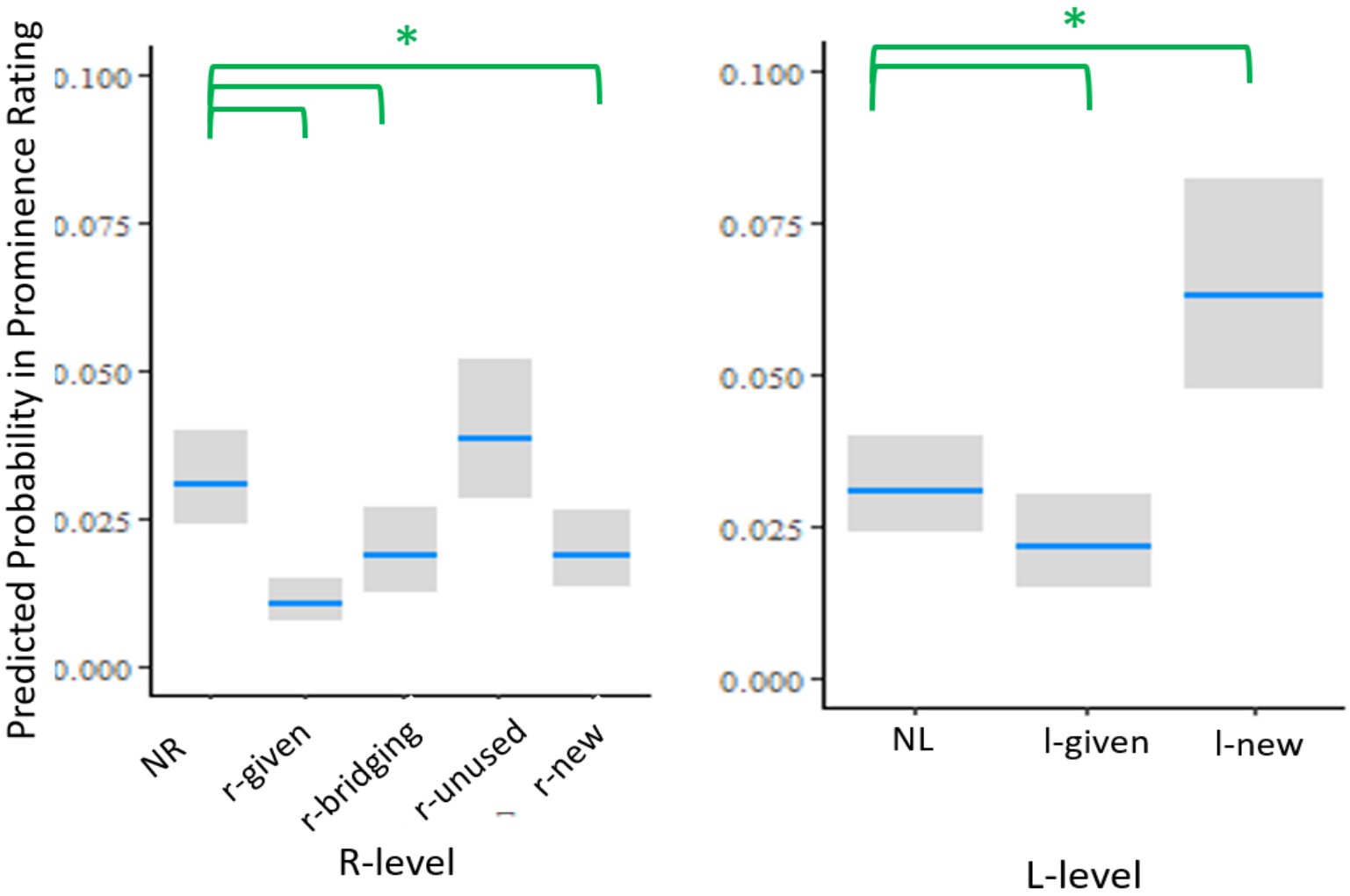
Results: Perceived Prominence and IS



Referential Givenness

- NR (Non-Referential): mostly function words, discourse markers, predicate expressions
- r-new

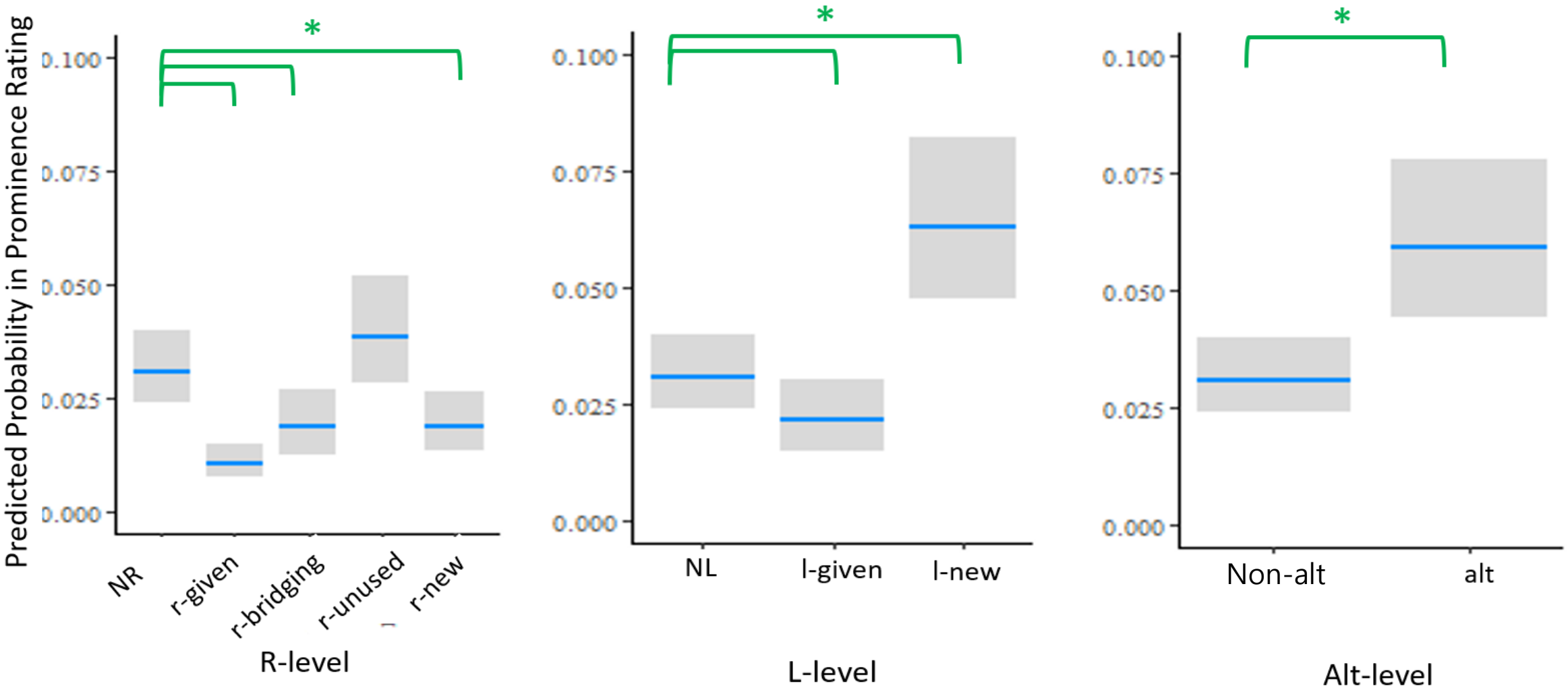
Results: Perceived Prominence and IS



Lexical Givenness

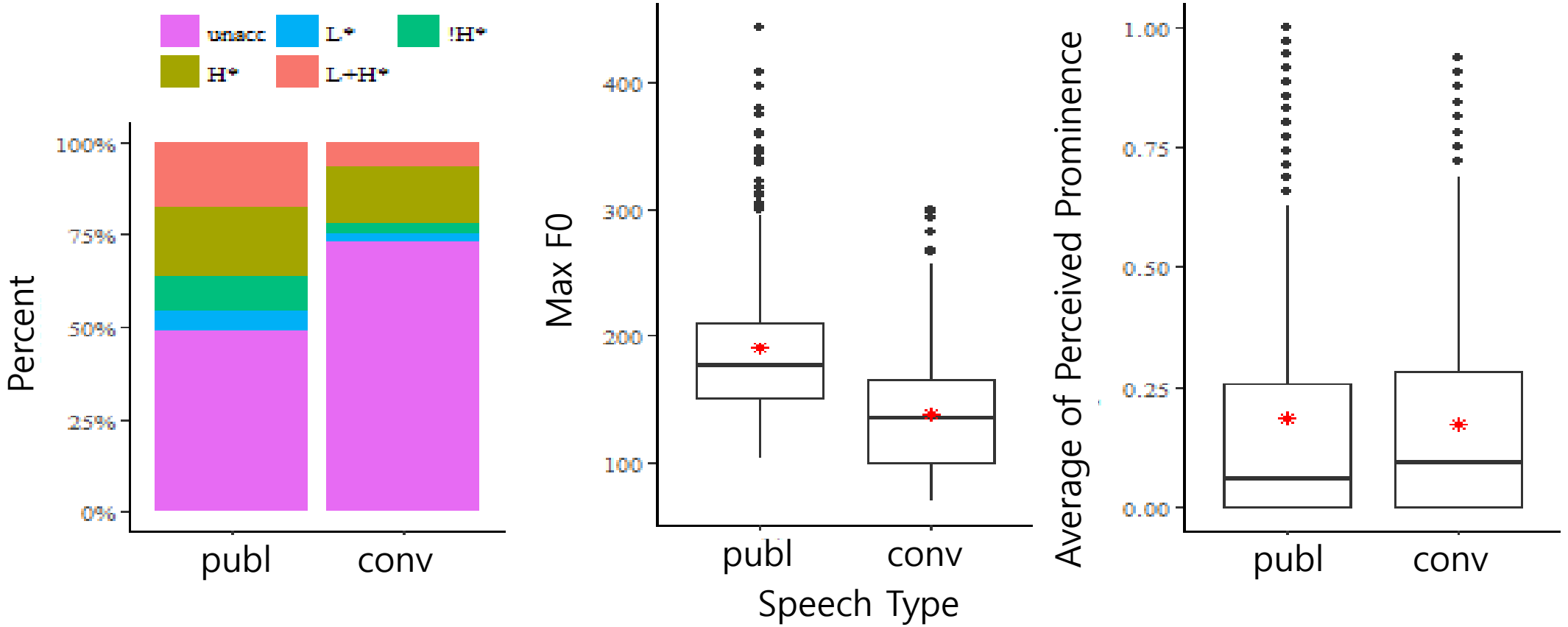
- NL (Non-Lexical): mostly function words, discourse markers, quantifiers

Results: Perceived Prominence and IS



Speech Style

- Comparison between the public speech (publ) and conversational speech (conv) from the Buckeye corpus (Cole et al., 2014)



Take-home messages

- The perception of prominence is influenced by **meaning-driven**.
- **Referential givenness** is differentiated from **lexical givenness** in American English.
- This study shows that **speech style** is also an important factor in the analysis of prosodic prominence.

Related research in progress

- Perception of prosody by **Korean learners of English**

- Prosody in the **mental representation** using…
  - **Statistical modelling** (e.g. Generalized Additive Model)
  - **Event-Related Potentials** (ERPs)