24.910 Topics in Linguistic Theory: Laboratory Phonology
Spring 2007

Effects of the lexicon and context on speech perception

Lexical Statistics

The pronunciation of words does not just depend on their phonological representations (features etc), also

- Prosody (phrasing, accentuation)
- Speech rate
- ‘Lexical’ statistics:
  - word frequency
  - neighborhood density
- Contextual predictability (cloze probability)

Lexical Statistics

- Word recognition is also affected by these properties
- It has been hypothesized that the production and perception effects are linked.
  - Words that are more difficult to recognize are pronounced more clearly.

Outline

- Review some effects of lexical statistics on word recognition.
- Present an analysis of these effects in terms of a Bayesian model of word recognition.
- Explore predictions concerning interactions between frequency/neighborhood density and contextual predictability.
  - These effects should be less important where contextual information is available.
- Next time: look at corresponding production effects, and general evidence for ‘listener-oriented behavior’ on the part of speakers.

Effects of lexical properties on word recognition

- Frequency: more frequent words are identified more rapidly and accurately (e.g. Goldinger et al 1996)
- Luce (1986) demonstrated that word frequency alone is not a very good predictor of difficulty in word recognition - neglects competition effects.
- Recognizing a word involves picking out that word from all of the words in the lexicon.
- This process of discrimination may be impeded where there are many words that are perceptually similar to the target word - lexical neighbors.
- High frequency facilitates the recognition of the target words, but high frequency neighbors impede recognition.

Neighborhood density/Relative frequency

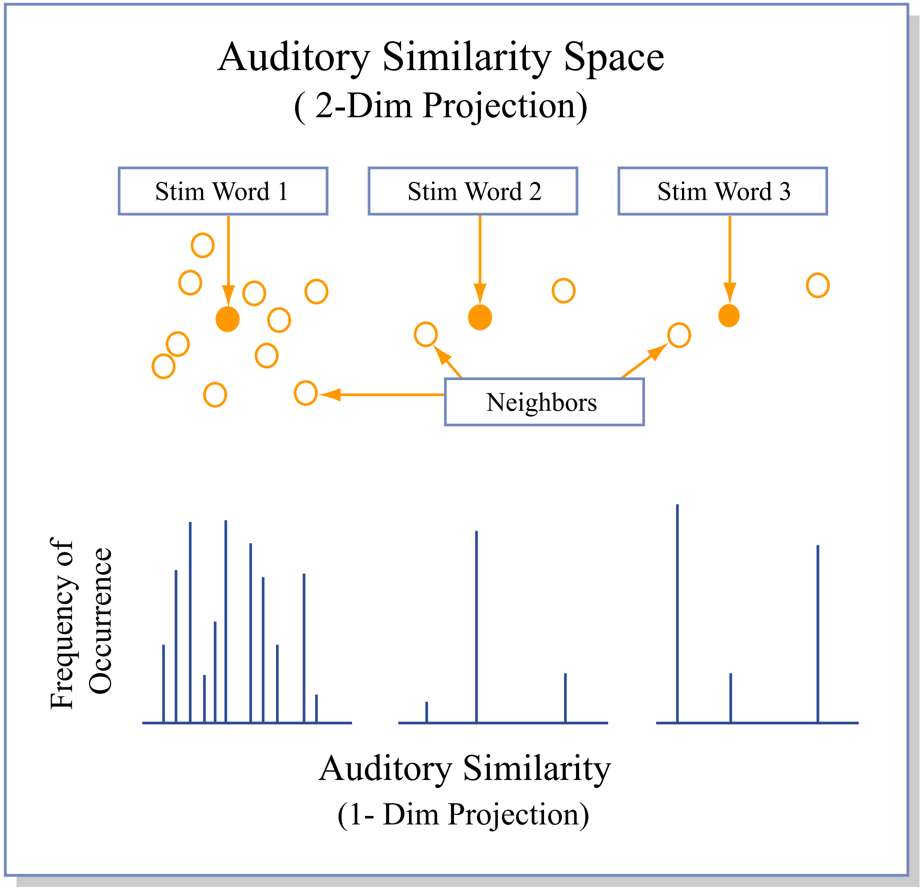

Figure by MIT OpenCourseWare. Adapted from Lindblom 1990.

Neighborhood density

- High density: cat: coat, at, scat, cap…
- Low density: choice: voice, chase
- Also matters how frequent those neighbors are

Vowel	High frequency / high density	High frequency / low density	Low frequency / high density	Low frequency / low density
a	got	dock	dot	mop
a	lock	rock	knock	sock
a	pot	top	cot	cop
æ	bad	bag	dad	dab
æ	sad	sang	fad	sag
æ	half	laugh	mash	rash
ɛ	get	death	debt	deaf
ɛ	bet	check	pet	pep
eɪ	save	gave	cage	bathe
eɪ	game	gain	dame	babe
eɪ	tape	shape	cake	nape

Figure by MIT OpenCourseWare.

Luce, Pisoni & Goldinger (1990)

- Tested effects of lexical neighborhood on speed and accuracy of identification of CVC words in noise.
- Neighborhood probability rule
  - p(stimulus word) is probability of correctly identifying the segments of the stimulus.
  - p(neighbor$_j$) is probability of misidentifying the stimulus as (having the segments of) neighbor *j*.

$$p(ID) = \frac{p(\text{stimulus word}) \times freq_s}{p(\text{stimulus word}) \times freq_s + \sum_{j=1}^{n} \left\{ p(neighbor_j) \times freq_j \right\}}$$

Luce, Pisoni & Goldinger (1990)

$$p(ID) = \frac{p(\text{stimulus word}) \times freq_s}{p(\text{stimulus word}) \times freq_s + \sum_{j=1}^{n} \left\{ p(neighbor_j) \times freq_j \right\}}$$

Predictions:

- words with higher frequency should be more accurately identified.
- Words with higher stimulus probability (made up of less confusable segments) should be more accurately identified.
- Words with more similar neighbors should be less accurately identified.
- Words with more high frequency neighbors should be less accurately identified.

Luce, Pisoni & Goldinger (1990)

- Stimuli: 400 CVC words divided into 8 classes, fully crossing:
  - High vs. low word frequency
  - High vs. low stimulus probability
  - High vs. low frequency-weighted neighborhood probability
- Words mixed with white noise (SNR +5dB) and presented to subjects for identification.
- Stimulus/neighbor probabilities were estimated from confusion matrices for CV and VC syllables in noise.
  - Assume confusion probability depends only on position.
  - $p(\text{kɪd}|\text{kæt}) = p(\text{k}_{ons}|\text{k}_{ons}) \times p(\text{ɪ}|\text{æ}) \times p(\text{d}_{coda}|\,\text{t}_{coda})$
  - p(∅|seg) and p(seg| ∅) were used to for CCVC, CV etc.
- Only familiar monosyllabic words were considered.

Results

- High stimulus probability words identified more accurately than low stimulus probability words.
- Words with high frequency-weighted neighborhood probabilities identified less accurately.
- High frequency words identified more accurately than low frequency words, but high freq words in dense neighborhoods identified less accurately than low freq words in sparse neighborhoods.

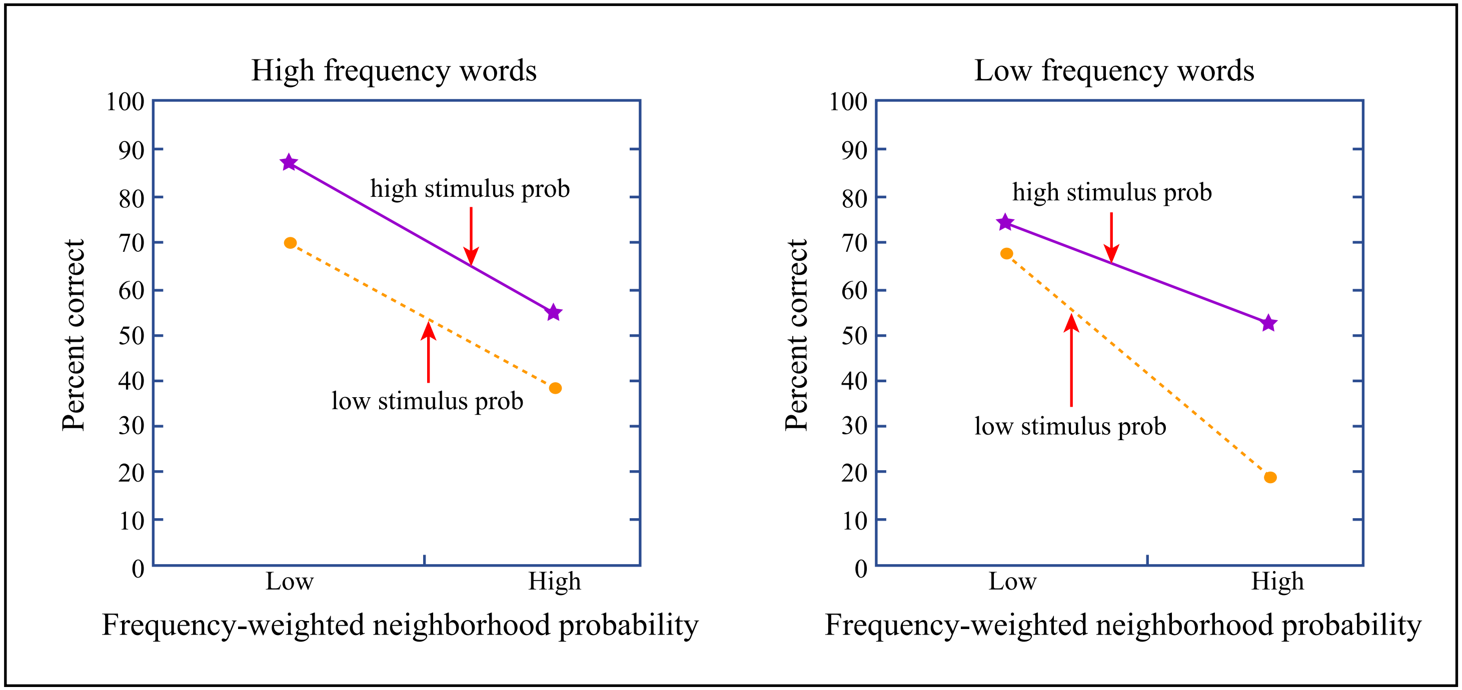

Figure by MIT OpenCourseWare.

Luce, Pisoni & Goldinger (1990)

- Lexical decision: Spoken word or nonword presented.
  - Subject must decide whether the stimulus is a word or not.
- Reaction time to nonword stimuli were slower where:
  - Mean frequency of neighbors is higher.
  - Density of neighborhood is higher.
  - No interaction.

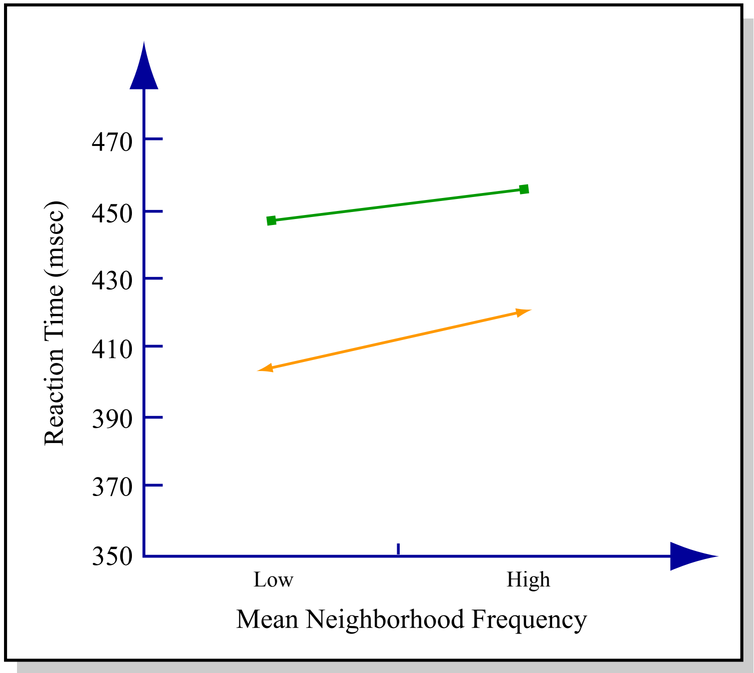

Figure by MIT OpenCourseWare. Adapted from Luce, P. A., D. B. Pisoni, and S. B. Goldinger. "Similarity Neighborhoods of Spoken Words." In *Cognitive Models of Speech Processing*. Edited by G. T. M. Altmann. Cambridge, MA: MIT Press, 1990, pp. 122-147.

- Here neighbors of a word are taken to be all words that can be created from that word by adding, deleting or changing one phone.
  - This operational definition is widely used.

A Bayesian model of word recognition

- The qualitative predictions of Luce's neighborhood probability rule can be reached based on a Bayesian model of word recognition (e.g. Jurafsky 1996, Norris 2006).
- Use Bayes Rule to combine signal-dependent and signal-independent evidence in word recognition.
- Probability of word *w* given signal-based evidence *E*:

(1) $$p(w \mid E) = \frac{p(E \mid w)\,p(w)}{p(E)}$$

prior probability of word

prior probability of evidence

(2) $$p(w \mid E) = \frac{p(E \mid w)\,p(w)}{\sum_{w_i \in lexicon} p(E \mid w_i)\,p(w_i)}$$

Bayes' Theorem

- Conditional probability:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} \qquad \Pr(B \mid A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

- Combine these equations:

$$\Pr(A \mid B)\Pr(B) = \Pr(A \cap B) = \Pr(B \mid A)\Pr(A)$$

- Divide by Pr(B), yieldng Bayes' Theorem:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$$

Application of Bayes' Theorem

- A medical test has a 95% chance of detecting a disease.
- The test has a 5% chance of yielding a positive result in the absence of the disease (false positive).
- 1 in 100 people has the disease.
- Suppose you have tested positive. What is the chance that you have the disease?

Application of Bayes' Theorem

$$P(Disease \mid Pos.Test) = \frac{P(Pos.Test \mid Disease)P(Disease)}{P(Pos.Test)}$$

P(Positive Test|Disease) = 0.95

P(Positive Test|¬Disease) = 0.05

P(Disease) = 0.01, P(¬Disease) = 0.99

P(Positive Test) = P(Pos.Test|Disease) × P(Disease) +
P(Pos.Test| ¬Disease) × P(¬Disease)
= 0.95 × 0.01 + 0.05 × 0.99 = 0.059

P(Disease|Pos.Test) = (0.95 × 0.01)/0.059 = 0.16

- Given the possibility of test error, we need to take prior probability into account.

A Bayesian model of the listener - word frequency

$$p(w \mid E) = \frac{p(E \mid w)\,p(w)}{\sum\limits_{w_i \in lexicon} p(E \mid w_i)\,p(w_i)}$$

- Evidence is accumulated over time. Listeners identify a stimulus as word *w* when that probability exceeds some threshold.
- Frequency: more frequent words are identified more rapidly and accurately (e.g. Goldinger et al 1996)
  - Higher frequency of *w* implies higher prior probability *p(w)*
  - Less bottom-up evidence required to reach a threshold probability that word is *w*.

    (Jurafsky 1996, Norris 2006, etc)

A Bayesian model of the listener - neighborhood density

$$p(w \mid E) = \frac{p(E \mid w)p(w)}{\sum_{w_i \in lexicon} p(E \mid w_i)p(w_i)}$$

- Neighborhood density: words from denser neighborhoods are identified more slowly and less accurately.
  - Neighbors of w are similar to w, so $p(E|w_i)$ is going to be relatively high where w_i is a neighbor.
  - So more neighbors and higher frequency neighbors increase the denominator above, reducing $p(w|E)$ (Jurafsky 1996).
  - NB standard calculation of neighborhood is an approximation (cf. Luce 1986).

A Bayesian model of the listener - context effects

- The Bayesian analysis implies that word frequency affects word recognition because it is a good basis for estimating prior probability of a word in the absence of any other constraint.
- But in general the prior probability of a word depends on context, e.g. discourse topic, previous words, syntactic structure.
- Ideal listener should incorporate these contextual effects into estimates of prior probabilities.

A Bayesian model of the listener - context effects

$$p(w \mid E,C) = \frac{p(E \mid w)\,p(w \mid C)}{\sum_{w_i \in lexicon} p(E \mid w_i)\,p(w_i \mid C)}$$

- The probability of a word depends on context *C*.
- increase in $p(w|C)$ reduces evidence needed for identification of *w*.
- Predictability effect: When words are more predictable from context they are:
  - more accurately identified (e.g. Boothroyd and Nittrouer 1988, Sommers and Danielson 1999).
  - Identified earlier in a gating task (Craig et al 1993).

Boothroyd & Nittrouer 1988

- Studied accuracy of word identification in nonsensical and meaningful sentences.
  - Zero predictability
    - Girls white car blink.
  - Low predictability
    - Ducks eat old tape.
  - High predictability
    - Most birds can fly.
- All words monosyllabic.
- Words from LP and HP sentences used in ZP sentences.

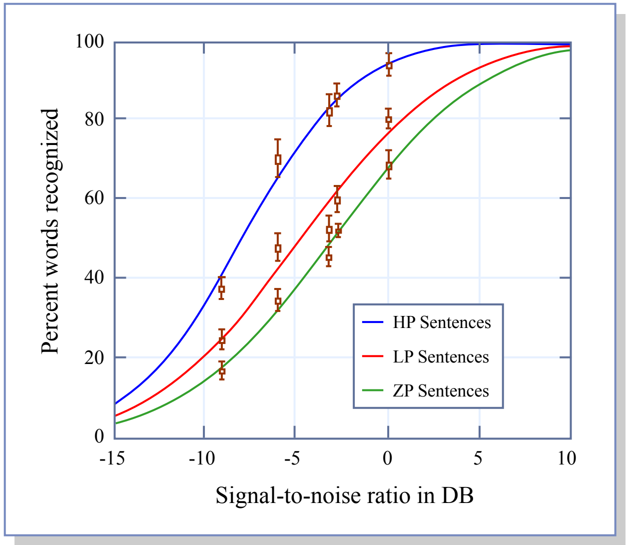

Figure by MIT OpenCourseWare. Adapted from Boothroyd, A., and S. Nittrouer. "Mathematical Treatment of Context Effects in Phoneme and Word Recognition." *Journal of the Acoustical Society of America* 84 (1988): 101-114.

A Bayesian model of the listener - context effects

$$p(w \mid E, C) = \frac{p(E \mid w)\, p(w \mid C)}{\sum_{w_i \in lexicon} p(E \mid w_i)\, p(w_i \mid C)}$$

The Bayesian model predicts interactions between predictability and frequency/neighborhood density:

- There is no word frequency term in the model - frequency only enters as an estimate of word probability $p(w|C)$ in the absence of contextual constraints.
- As context raises prior probability of w, the effect of competition from neighbors should be reduced.
  - $p(w|C)$ increases, most $p(w_i \neq w|C)$ decrease.
- As contextual constraint increases, the effects of word frequency and neighborhood density on word recognition should decrease.

Interactions between context and lexical statistics

- As contextual constraint increases, the effects of word frequency and neighborhood density on word recognition should decrease.
  - implies reduced importance for frequency per se for running speech (same for frequency of neighbors).

Interactions between context and lexical statistics

Frequency/Context:

- Grosjean & Itzler (1984): effect of frequency on the isolation point of gated words is reduced where words are more predictable from context (almost to zero in the most constraining contexts).
- Van Petten and Kutas (1990): ERP study of silent reading - less frequent words were associated with larger N400s early in sentences, but the frequency effect disappears later in a sentence, as semantic and syntactic constraints accumulate (also Dambacher et al 2006).
  - 'frequency does not play a mandatory role in word recognition but can be superseded by the contextual constraint provided by a sentence'

Interactions between context and lexical properties: Neighborhood density/Context

Sommers and Danielson (1999):

- Auditory word identification task
  - Isolated words.
  - Final words in sentences:
    - Low predictability: ‘She was thinking about the **path**’.
    - High predictability: ‘She was walking along the **path**’.
  - Words had high (28) or low (9.1) neighborhood density (‘hard’ vs. ‘easy’).
    - Matched for frequency.
- Two speakers, 22 listeners.
- Materials presented in noise.

Sommers & Danielson (1999)

Results

- Significant differences in identification accuracy across the three contexts.
- Significantly lower accuracy for words from dense neighborhoods.
- Effect of neigborhood density is reduced in High Predictability contexts (significant interaction Density × Context).

Percent correct

Context	Easy		Hard	
	M	SD	M	SD
Single Word	78.7	10.2	62.8	14.3
Low Predictability	84.4	9.1	69.7	11.4
High Predictability	92.1	4.3	84.4	6.7

Figure by MIT OpenCourseWare. Adapted from Sommers, M. S., and S. M. Danielson. "Inhibitory Processes and Spoken Word Recognition in Young and Older Adults: The Interaction of Lexical Competition and Semantic Context." *Psychology and Aging* 14 (1999): 458-472.

Interactions between context and lexical properties: Neighborhood density/Context

- Sommers, Kirk and Pisoni (1997): difference in accuracy of identification of ‘hard’ and ‘easy’ words disappeared where subjects had to pick from a closed set of words.

- Bayesian model provides an accurate qualitative characterization of the effects of frequency, neighborhood density and contextual predictability on word recognition performance.
- Two basic factors:
  – Competition within the lexicon.
  – Predictability of target and competitors.
    - Frequency is an estimate of probability in the absence of context.

References

- Aylett, M. and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *JASA* 119:3048-3058.
- Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *JASA* 113, 1001-1023.
- Billerey-Mosier, R. (2000). Lexical effects on the phonetic realization of English segments. *UCLA Working Papers in Phonetics* 100.
- Boothroyd and Nittrouer (1988). Mathematical treatment of context effects in phoneme and word recognition. *JASA* 84(1):101.
- Dambacher, M., Kliegl, R., Hofmann, M., and Jacobs, A.M. (2006). Frequency and predictability effects on event-related potentials during reading, *Brain Research* 1084, 89-103.
- Goldinger, S.D., Pisoni, D.B., & Luce, P.A. (1996). Speech perception and spoken word recognition: Research and theory. In N.J. Lass (Ed.), *Principles of Experimental Phonetics.* St. Louis: Mosby. 277-327.
- Griffin, Z.M., and Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language* 38, 313-338.

References

- Grosjean, F., & Itzler, J. (1984). Can semantic constraint reduce the role of word frequency during spoken word recognition? *Bulletin of the Psychonomic Society* 22, 180-182.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20, 137-194
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W.J. Hardcastle and A. Marchal (eds.) *Speech Production and Speech Modeling*. Kluwer: Dordrecht
- Luce, P., and Pisoni, D. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19. 1-36.
- Munson, B. (2004) Lexical access, lexical representation, and vowel production. To appear in J. Cole and J. I. Hualde (eds.), *Papers in Laboratory Phonology IX*. Mouton de Gruyter.
- Munson, B., and Solomon, N.P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47, 1048-1058.
- Norris, D. (in press) The Bayesian Reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review.*

References

- Pierrehumbert, J.B. (2002). Word-specific phonetics. In C. Gussenhoven and N. Warner (eds.) *Papers in Laboratory Phonology VII*, Mouton de Gruyter, New York, 101-139.
- Scarborough, R. (2003). The word-level specificity of lexical confusability effects. Poster presented at the 146th Meeting of the Acoustical Society of America, Austin, TX.
- Sommers, M.S., and Danielson, S.M. (1999). Inhibitory processes and spoken word recognition in young and older adults: the interaction of lexical competition and semantic context. *Psychology and Aging* 14, 458-472
- Sommers, M., Kirk, K., Pisoni, D. (1997). Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I. The effects of response format. *Ear and Hearing* 18, 89–99.
- Van Petten, C. and Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition* 18, 380-393.
- Van Son, R.J.J.H. and Pols, L.C.W. (2003a). Information Structure and Efficiency in Speech Production", Proceedings of EUROSPEECH2003, Geneva, Switzerland, 769-772.

References

- Van Son, R.J.J.H. and Pols, L.C.W. (2003b). An Acoustic Model of Communicative Efficiency in Consonants and Vowels taking into Account Context Distinctiveness". Proceedings of ICPhS, Barcelona, Spain, 2141-2144.
- Vitevitch, M. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28, 735-747.
- Wright, R. (2004) Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, and R. Temple (eds.), *Papers in Laboratory Phonology VI.* Cambridge: CUP.

A Bayesian model of the listener - context effects

- The Bayesian analysis implies that word frequency affects word recognition because it is a good basis for estimating prior probability of a word in the absence of any other constraint.
- But in general the prior probability of a word depends on context, e.g. discourse topic, previous words, syntactic structure.
- Ideal listener should incorporate these contextual effects into estimates of prior probabilities.