24.910 Topics in Linguistic Theory: Laboratory Phonology
Spring 2007

Effects of the lexicon and context on speech production

Word recognition

The speed an accuracy of word recognition depends on:

- Word frequency
- Neighborhood density
  - and frequency of neighbors
- Contextual predictability

- Speech production is also affected by these factors.

Effects of lexical statistics on production

- Wright (2004) found that neighborhood density/relative frequency affects pronunciation of isolated words.
- ‘Hard words’ - low frequency, high neighborhood density
- ‘Easy words’ - high frequency, low neighborhood density
- Vowels in hard words are more dispersed from each other in F1*F2 space than vowels in easy words.

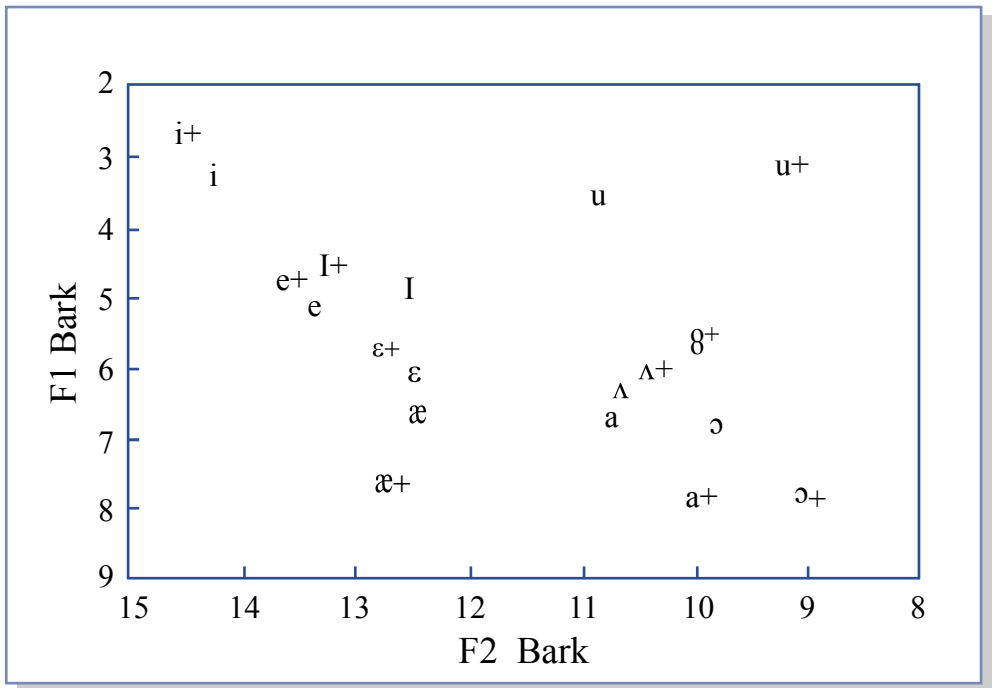

Figure by MIT OpenCourseWare.

Listener-Oriented Speakers

- It has been hypothesized that the production and perception effects are linked:
  - Words that are more difficult to recognize are pronounced more clearly.

Production and perception

Broad outlines of an explanatory model of the effects of frequency, neighborhood density etc on production (Wright 2004, Scarborough 2004, 2006):

- Speaker wishes to be understood.
- Speaker wishes to minimize the effort involved in speech production.
- Reduced effort tends to lead to reduced identifiability of words.
- Optimal strategy: reduce effort more where clarity is less important, i.e. where top-down evidence makes it easier for a listener to identify a word - Hyper- & Hypoarticulation theory (Lindblom 1990).

Are speakers altruistic or selfish?

- Implication of H&H account of pronunciation variability: Speakers estimate listener difficulty moment to moment and adjust clarity of speech accordingly.
  - Speaker has a model of the listener (altruistic speaker).
- Alternative line of analysis: Speakers are selfish.
  - Speakers do not track listener difficulty.
  - Pronunciation variation is related to speaker difficulty with lexical access for production:
    - Slower lexical access results in clearer speech.
  - To the extent that ease of lexical access is similar for speaker and hearer, similar results are predicted (but there are differences).
  - But why does slow lexical access result in clearer speech?

Frequency and Neighborhood Density: Munson & Solomon (2004)

- Point out that Wright (a) confounded frequency and neighborhood density, (b) didn't measure duration, so we can't be sure if neighborhood density affects vowel formants directly or via vowel duration.

Vowel	Lexically easy	Lexically difficult
a	job	cod
a	shop	cot
a	wash	knob
a	watch	wad
æ	gas	hack
æ	jack	hash
æ	path	pat
ɪ	give	hick
ɪ	ship	kin
ɪ	thing	kit
i	peace	bead
i	teeth	weed
o	both	goat
o	vote	moat
u	food	hoop

Figure by MIT OpenCourseWare.

Experiment 1

- Words read in isolation by 10 subjects.
- Vowel space is expanded in Hard words (mean Euclidian distance of vowels from the mean F1, F2 of all vowels).
- Vowels were shorter in hard words (222 ms vs. 232 ms).

Experiment 2

- 4 classes of words, crossing:
  - high vs. low frequency
  - high vs. low neighborhood density
- 20 words /class
- Duration:
  - High frequency words had shorter vowels (205 ms vs. 211 ms)
  - No effect of neighborhood density.
- Vowel space expansion:
  - Less expanded in high frequency words.
  - More expanded in high neighborhood density words.
  - No significant interaction.

Selfish speakers and neighborhood density

- The H&H account: Words from dense neighborhoods/low frequency words are pronounced more clearly because the speaker knows they are likely to be more difficult for listeners to recognize.
- Speaker-oriented account (e.g. Pierrehumbert 2002):
  - Speakers have to perform lexical access in speech production.
  - Hypothesize that high neighborhood density impedes lexical access in production.
  - Slower lexical access results in clearer pronunciation.

Selfish speakers and neighborhood density

- But high neighborhood density can actually speed lexical access in production: Pictures are named more quickly when their names are in dense neighborhoods (Vitevitch 2002).
- Lexical access in production starts from meaning, so there is no problem of competition based on phonetic similarity.
- Vitevitch offers two explanations for the facilitatory effect of dense neighborhoods:
- In an ‘interactive activation model’: activation spreads between phonetically similar words. In dense neighborhoods more activation ‘reverberates’ back to the target word.
- OR: words in dense neighborhoods generally involve more common sound sequences - perhaps the motor plans for frequent sound sequences are easier to access/assemble.

Lexical access and neighborhood density

- Munson (2004) directly tested the ‘lexical access’ account of neighborhood density effects:
- Four classes of words crossing:
  - High vs. low frequency
  - High vs. low neighborhood density
- Subjects read words in two conditions:
  - Read word immediately on presentation.
  - Wait 1000 ms after presentation before speaking word.
- Assumption: 1000 ms delay gives speakers plenty of time to complete lexical access, so difficulty with lexical access should not affect pronunciation at this time lag.

Munson (2004) results

- In both conditions vowels were more dispersed in words from dense neighborhoods.
  - suggests effect of neighborhood density is not due to speaker difficulty with lexical access.
- The effect of frequency depends on condition:
  - No delay: vowels more dispersed in low frequency words.
  - Delay: effect of frequency not significant.

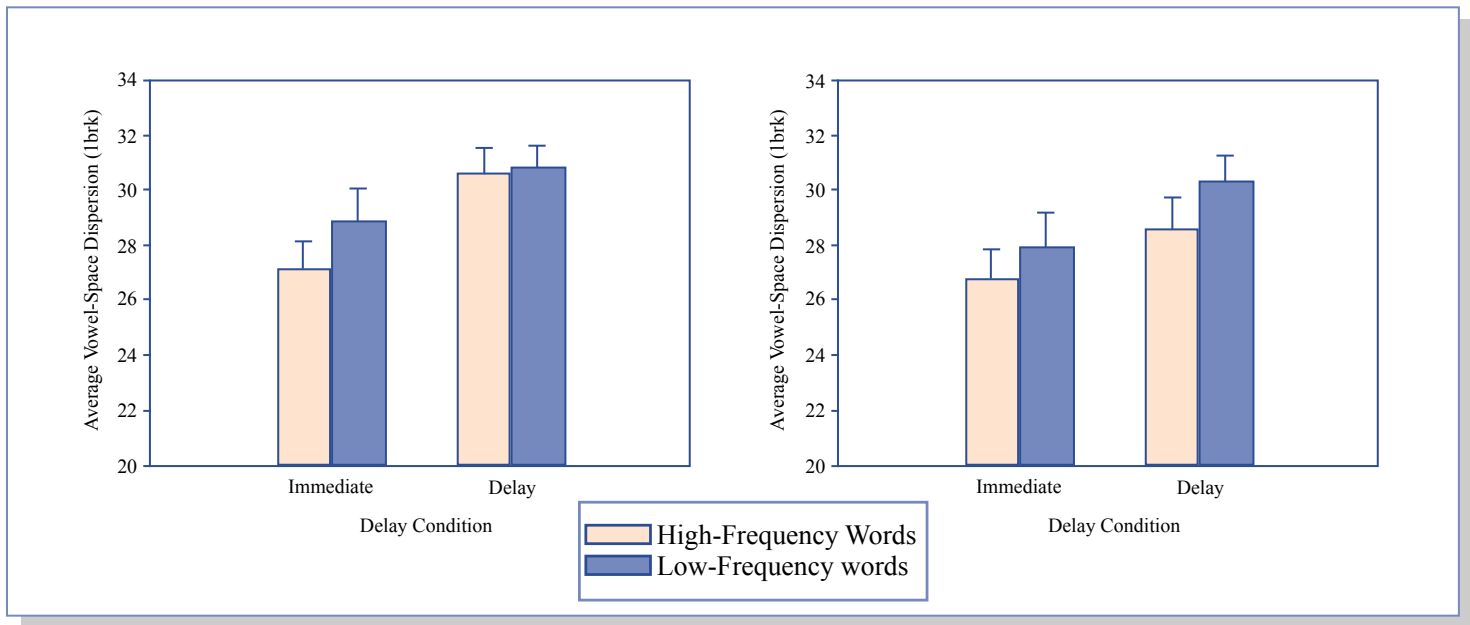

Figure by MIT OpenCourseWare.

Contextual predictability and speech production

- H&H theory predicts that speaker clarity should generally mirror listener difficulty with lexical access.
- Given the evidence above concerning word recognition performance, we should expect to find:
  1. Reduced clarity where a word is predictable from context.
  2. Reduced effects of frequency/neighborhood density in contexts where a word is more predictable.
- Some evidence for (1), e.g Lieberman (1963), Hunnicutt (1985), Fowler & Housum (1987), Bell et al (2003).
- Only one inconclusive test of (2).

Contextual predictability and speech production

- Lieberman (1963):
  - The word you will hear is nine.
  - A stitch in time saves nine.
- Words excised from predictable contexts are less accurately identified in noise.
- Hunnicutt (1985) replicated and extended this result.
  - Assessed predictability of words by Cloze procedure: Present sentences with word left out. Ask subjects to guess the missing word. Cloze probability = proportion of correct guesses.
- No investigation of the acoustic bases of these effects.

Repetition and speech production

- Fowler & Housum (1997) examined the pronunciation and perception of first and second uses of words in a monologue (Garrison Keillor - 35 pairs of words) and some short news interviews (45 pairs).
- Second mentions should generally be more predictable.
- Very high frequency words excluded (function words).
- Measured:
  - Word duration.
  - Peak amplitude.
  - f0 of stressed vowel.

Repetition and speech production

- New words were longer, louder and had higher f0.
  - Only the duration effect is statistically reliable across speakers.

	Duration (ms)		f_0 (Hz)		Amplitude (volts)	
	New	Old	New	Old	New	Old
Keillor	562	492	119	110	1.12	1.03
Others	436	395	135	134	1.92	1.77

Figure by MIT OpenCourseWare.

- A subsequent experiment established that old words were, on average, more predictable in a Cloze task:
  - New 18.3% correct vs. Old 31.1% correct

Repetition and speech production

- Words from the first experiment were excised and presented for identification by subjects (+ confidence rating /5)
  - Both versions of each word were heard in separate blocks.
  - Order or pairs counter-balanced across subjects.
- Repeated words are less accurately identified.
- Accuracy correlated with duration difference (new-old).
- Words identified more accurately on second presentation.
  - repetition facilitates recognition (i.e. could offset reduction).

Block of test	Occurrence in monolog	
	1st	2nd
	Percentage errors	
1	11.6	16.2
2	8.8	12.4
	Confidence judgments	
1	4.47	4.26
2	4.69	4.49

Self-centered behavior in production of repetitions

- Bard et al (2000) examined shortening of repeated words in spontaneous dialogues produced in the context of a ‘map task’.
- Two participants each have a map, one showing a route. Not all landmarks are shown on both maps.
- Subject with the route map instructs the second subject on how to reproduce the route on his/her map.
- Elicits spontaneous speech, but with multiple mentions of items marked on the maps.
- Basic set-up: look for effects of repetition on intelligibility of words.
- Twist: subjects repeat the task with two partners. Will they produce clear, longer pronunciations on first mentions to a new listener?

Self-centered behavior in production of repetitions

- Each subject lead two partners through the same map task.
- Key words were excised from the recorded dialogues and presented in noise for identification in a follow-up experiment.
- Durations of words were also measured.

Self-centered behavior in production of repetitions

- Intelligibility and duration of first mentions of landmarks were reduced in second trial, even though they were new to the second listener.
- Speakers seem to act as if an item is old if it is old to them, even if it is new to the listener (‘egocentric’).

Face condition	Form		
	Trial 1	Trial 2	Citation
	Intelligibility		
Screened	0.746 (0.072)	0.636 (0.182)	0.818
Visible	0.578 (0.230)	0.693 (0.115)	0.808
	Duration		
Screened	0.703 (0.606)	0.659 (0.650)	1.309
Visible	0.646 (0.655)	0.588 (0.713)	1.301

Self-centered behavior in production of repetitions

- Bard et al's results might be attributed to effects of lexical access (priming).
- Bard et al actually argue that speakers are incompetent altruists: tracking listener needs is simply too demanding, so speakers usually make the simplifying assumption that listeners know what they know.
- A further complication: Bard et al (2000) did not compare repeated interactions with the same vs. different listeners. The second listener was always new.
- Gregory (2001) had speakers tell a story twice, either to the same listener or to two different listeners.
  - Measured durations of repeated referring expressions (e.g. names).
  - When hearer changes for second narration, first use in second narration is longer than first use in narration 1 (n.s.)
  - When hearer is the same for both narrations, first use in narration 2 is shorter than first use in narration 1.
  - Indicates that speakers can take listener knowledge into account.

Listener-oriented behavior in speech production

- While a variety of the effects predicted by the H&H model have been observed (neighborhood density, frequency, predictability), there are alternative ‘selfish’ explanations for most of them.
- H&H theory offers a simple, unified account of these phenomena, the speaker-oriented analysis are more ad hoc (e.g. why should slower lexical access result in clearer speech? Lexical access cannot explain neighborhood density effects, etc.)
- It is clear that H&H theory needs to be supplemented by an account of how well speakers track listener needs.

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.