Predictability and Phonetic Attention:
How Context Shapes Exemplars

Jonathan Manker
PHREND
University of California, Santa Cruz
September 24, 2016
Overview of Presentation

• I. Background
  1) Modes of Listening
  2) Exemplar Theory

• II. Experiment I: Listening to Predictable and Unpredictable Words

• III. Experiment II: Imitation of Predictable and Unpredictable Words

• IV. Discussion and Conclusion
Background I: Modes of Listening

• Research has shown low-level phonetic information is used alongside higher-level contextual information in speech recognition.
• Cole & Jakimik (1980): “Words are recognized through the interaction of sound and knowledge.”
• Words excised from sentence contexts can be phonetically ambiguous or even unrecognizable (coarticulation, reduction, etc.)
• Phoneme restoration (Warren 1970): speakers fail to notice [s] replaced with a cough
• Marslen-Wilson & Welsh (1978): speakers more likely to notice onset rather than coda speech errors ([n] replacing /m/ in ‘made’ vs. ‘time’)


‘What’ vs. ‘How’ Mode

• Lindblom et al. (1995):
  • ‘what’ mode: more commonly used in speech, listeners focuses on extracting meaning
  • ‘how’ mode: listener focuses on details of sounds, pronunciation

• relevant in sound change: suggests Ohala’s hypocorrection could result from listening in ‘how’ mode, failing to normalize expected coarticulation
Listening Modes: Neurological Evidence

• apply *dorsal* and *ventral* streams to auditory processing
• *Ventral stream*:
  • lower part of brain, bilateral
  • mapping between abstracted sound and meaning
• *Dorsal stream*:
  • upper portions of brain, left-hemisphere oriented
  • active in sublexical processing such as phoneme identification, rhyming tasks
  • involved in motor planning; suggests perception-production link
• Ventral analogous to ‘what’ mode; dorsal to ‘how’ mode
Background: Exemplar Theory

• Exemplar-based theories of speech (Johnson 1997, Pierrehumbert 2002) propose that individual instances of words are stored in memory.

• Contrasted with theories only allowing abstract representations of words

• Speakers may use all the available phonetic details to identify words, rather than normalizing all speech to abstracted forms

• More recent work has suggested a need for both exemplars and abstract representations (Pierrehumbert 2002, Goldinger 2007).
Exemplars and Higher-Level Knowledge

• Goldinger (2007): “[E]ach stored exemplar is actually a product of perceptual input combined with prior knowledge…”

• Maye (2007): Exemplars may be shaped by different levels of attention towards different cues, based on subjects’ L1

• Pierrehumbert (2006): Speakers attuned to more informative socio-indexical features
The Current Study

• Cole & Jakimik’s observations demonstrate the effect of context on speech recognition and perception

• Does context then mediate the activation of ‘what’/ventral vs. ‘how’/dorsal listening modes?

• Hypotheses:

  • More attention to phonetic detail should occur with less context.
  • Predictable speech may be stored as less precise exemplars, more heavily influenced by higher level linguistic information.
Experiment I: Listening to Predictable and Unpredictable Speech

• Target question: Do listeners store or process more detail for unpredictable words, and less for predictable words?

• This experiment will focus on speech perception alone, to isolate the role of the listener in attending to phonetic detail.

• Subjects will hear words in sentences, in either a predictable or unpredictable context; the word will be repeated after a pause and listeners will be asked to compare it to the original. Is the same?

• Prediction: Subjects will more notice more phonetic details of unpredictable words and will make more errors in comparing predictable words.
Experiment I: Stimuli

• Subjects are presented with a total of 80 recorded sentences:
  • Of these, 30 are target sentences including:
    • 15 with target words in predictable contexts
    • 15 with different target words in unpredictable contexts

• Target words: All target sentences end in a target word, beginning with /k/, two syllables, and initial stress.

• Predictable Context:
  • “The pioneers made log ______,” or “Pennies are made out of ______.”

• Unpredictable Context:
  • “Joe turned and saw the ______,” or “The next word is ______.”

• Predictability of words confirmed in Mechanical Turk survey: 30 subjects were asked to give one guess about blank word. All target words were correctly guessed by between 9/30 and 30/30 participants.
Experiment I: Stimuli and tasks

• The same recordings of the target words are heard in both predictability contexts (copied from predictable, pasted into unpredictable; pronunciations are identical, only difference is surrounding context)

• The subjects will hear these target sentences, with either predictable or unpredictable targets, followed by a pause and some static, then with the target word repeated.

• The target word is copied from the sentence, and is repeated either identically, or is modified to have longer VOT (at least 100 ms) and higher initial pitch (20 Hz higher than model).
• Examples: Pennies are made out of **copper** … [STATIC] … **copper**
  The next word is **copper** … [STATIC] … **copper**

• Subjects are asked to judge if the repeated word sounds the same as when it occurred in its sentence.

• Prediction: Subjects will be more accurate in their comparisons with *unpredictable* words and less so with predictable words (more errors)
The remaining 50/80 sentences are fillers of two types:

- 25 audio comparison sentences (just like the targets), except the target / repeated word is not the last word in the sentence, and does not begin with /k/
- 25 “content” questions such that:
  - The listener simply hears “The professor taught linguistics for sixty years...” followed by “How long did the professor teach linguistics?
  - These were used in order to keep subjects listening both for meaning (“what?”) and sound (“how?”)
Subject Groups

• 80 subjects participated via Amazon Mechanical Turk (with the experiment run using SurveyGizmo)

• 4 groups with 20 subjects each:
  • 2 groups were presented with 30/60 of the target words, while the other 2 groups were presented with the other half of the words
  • Within these groups, the targets were counterbalanced for predictability, so target words were only heard once by each subject, either predictable or unpredictable

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td>heard e.g. ‘cabins’ in pred. context, ‘copper’ in unpred. context</td>
<td>heard e.g. ‘canvas’ in pred. context, ‘college’ in unpred. context</td>
</tr>
<tr>
<td>Group A’</td>
<td>Group B’</td>
</tr>
<tr>
<td>heard e.g. ‘cabins’ in unpred. ‘copper’ in pred. context</td>
<td>heard e.g. ‘canvas’ in unpred. context, ‘college’ in pred. context</td>
</tr>
</tbody>
</table>
Results

• The target data from this experiment was all categorical Y/N responses (did the word sound the same when repeated?)
• Over 80 subjects and 2400 total responses (for target stimuli)
• Of those 2400, there were 535 errors (22.3% error rate)

• Subjects made significantly more errors in comparing predictable words \((p = 0.012)\) following my hypothesis.
• Errors for predictable words: 291 (54.4%)
• Errors for unpredictable words: 244 (45.6%)
### Analysis of Deviance Table (Type II Wald chisquare tests)

**Response:** response ~ answer * group * predictability * + (1|subject) + (1|word)

<table>
<thead>
<tr>
<th></th>
<th>Chisq</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>answer</td>
<td>5.4117</td>
<td>1</td>
<td>0.02000*</td>
</tr>
<tr>
<td>group</td>
<td>2.1484</td>
<td>1</td>
<td>0.14272</td>
</tr>
<tr>
<td>predictability</td>
<td>6.2991</td>
<td>1</td>
<td>0.01208*</td>
</tr>
<tr>
<td>answer:group</td>
<td>1.7661</td>
<td>1</td>
<td>0.18387</td>
</tr>
<tr>
<td>answer:predictability</td>
<td>0.1064</td>
<td>1</td>
<td>0.74424</td>
</tr>
<tr>
<td>group:predictability</td>
<td>0.9430</td>
<td>1</td>
<td>0.33152</td>
</tr>
<tr>
<td>answer:group:predictability</td>
<td>1.6281</td>
<td>1</td>
<td>0.20196</td>
</tr>
</tbody>
</table>

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

![Pie chart showing the distribution of predictable and unpredictable false alarms.](chart.png)
Experiment II: Imitation of Predictable and Unpredictable Speech

• The first experiment established that predictability modulates attention to phonetic details
• Thus, contextual information affects speech perception
• Could this phenomenon be relevant for sound change?
• Experiment II will examine the listener turned speaker--- how perception influences production
• Phonetic accommodation paradigm will be used to understand what phonetic details are perceived and how those affect production
Phonetic Accommodation

• phenomenon whereby characteristics (phonetic, syntactic, etc.) of one’s speech are influenced through perceiving the speech of others.

• stored word exemplars determine pronunciation

• new exemplars influence new productions

• Earliest observation in lab setting, Goldinger (1998):
  • baseline word reading to post-stimulus
  • immediate vs. delayed shadowing
  • word frequency, # of repetitions
  • more impressionistic AXB format for comparing to model

• Quantitative measurements of specific manipulated features:
  • Shockley, Sabadini & Fowler (2004), Nielsen (2011): VOT
  • Tilsen (2009): vowel quality
Phonetic Accommodation, cont’d

• Other features of the phenomenon:
  • automatic / subconscious? (Goldinger 1998, Lewandowski 2012)
  • Specificity of accommodated features: Nielsen (2011)
  • Abstractness / generalizability: Nielsen (2011)
• Social factors:
  • Divergence/convergence: Babel (2010), Babel (2012)
  • Gender and conversational roles: Pardo (2006)
• Thus, phonetic accommodation is a useful tool for understanding what was perceived and how this influences production

• Hypothesis: More phonetic accommodation will be observed for unpredictable speech
Stimuli

• The same target words and sentences were also used for experiment II.

• **Predictable Context:** “The pioneers made log _____,” or “Pennies are made out of _____.”

• **Unpredictable Context:** “Joe turned and saw the _____,” or “The next word is _____.”

• In all target sentences, the final /k/ initial word was digitally enhanced (VOT was approximately doubled to at least 100 ms, pitch of the first syllable was raised by about 20 Hz).

• Once again, the target words were excised from their pronunciations in the predictable context into the unpredictable context; thus the same recording of the same word was heard in both conditions (undoing any possible hyperarticulation of unpredictable words).
Tasks

• The experiment consisted of a single block of 100 sentences in random order:
  • 30 with target predictable words
  • 30 with target *different* unpredictable words
  • 40 filler sentences (no /k/ final words, no digital enhancement)

• No target word was ever heard twice in the experiment (no priming effect possible)
Subject Groups

• The experiment consisted of four groups of 10 subjects each (40 total).

• Two instructional conditions:
  • Condition I: No instruction to imitation (20 subjects)
  • Condition II: Told to “sound more like the model in some way” (20 subjects)

• Groups were counterbalanced for target word context:
  • Group A heard 30 target predictable set A, 30 target unpredictable set B, and 40 fillers
  • Group B heard 30 target predictable set B, 30 target predictable set A, and the same 40 fillers.
  • (Again, each subject heard each target word only once, in either a predictable or unpredictable context depending on the group)
<table>
<thead>
<tr>
<th></th>
<th>No instruction to imitate</th>
<th>Told to imitate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30 unpredictable: “The first thing Mary saw was the coffins.”</td>
<td>30 unpredictable: “The first thing Mary saw was the coffins.”</td>
</tr>
<tr>
<td></td>
<td>40 fillers</td>
<td>40 fillers</td>
</tr>
<tr>
<td><strong>Group A’ – counterbalanced</strong></td>
<td>30 predictable: “The vampires are sleeping in coffins.”</td>
<td>30 predictable: “The vampires are sleeping in coffins.”</td>
</tr>
<tr>
<td><strong>(reverse predictability)</strong></td>
<td>30 unpredictable: “Joe turned and saw the cabins.”</td>
<td>30 unpredictable: “Joe turned and saw the cabins.”</td>
</tr>
<tr>
<td></td>
<td>40 fillers</td>
<td>40 fillers</td>
</tr>
</tbody>
</table>
Results

• With only one block (no baseline or post-exposure reading, which would have primed words in listening task), our question was:

*Is the subjects feature X (VOT, pitch, etc.) more similar to the model’s in the predictable or unpredictable context?*
Results: VOT

- $\text{votdiff} = \text{vot}_{\text{subj}} - \text{vot}_{\text{model}}$
- $\text{relvotdiff} = (\text{vot}_{\text{subj}} / \text{vl}_{\text{subj}} - \text{vot}_{\text{mod}} / \text{vl}_{\text{mod}})$

<table>
<thead>
<tr>
<th></th>
<th>NO INSTRUCTION TO IMITATE</th>
<th>TOLD TO IMITATE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predictable words</td>
<td>Unpredictable words</td>
</tr>
<tr>
<td>VOTDIFF</td>
<td>mean -59.0 ms</td>
<td>-54.1 ms</td>
</tr>
<tr>
<td></td>
<td>median -58.6 ms</td>
<td>-52.0 ms</td>
</tr>
<tr>
<td>RELVOTDIFF</td>
<td>mean -56.98%</td>
<td>-50.15%</td>
</tr>
<tr>
<td></td>
<td>median -55.49%</td>
<td>-49.09%</td>
</tr>
</tbody>
</table>

- Higher values (closer to zero) indicate VOT closer to the model. Subjects were closer to the model when told to imitate, but more importantly were closer to the model for unpredictable words.
Mixed effects model over all subjects reveals predictability and order within experiment are significant predictors of VOT.

Analysis of Deviance Table (Type II Wald chisquare tests)

Response: RelVotDiff ~ predictability * group * order * gender + (1|subject) + (1|word)

<table>
<thead>
<tr>
<th></th>
<th>NO INSTRUCTION TO IMITATE</th>
<th>TOLD TO IMITATE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chisq  Df Pr(&gt;Chisq)</td>
<td>Chisq  Df Pr(&gt;Chisq)</td>
</tr>
<tr>
<td>predictability</td>
<td>24.6517 1 6.868e-07 ***</td>
<td>1.3089 1 0.25260</td>
</tr>
<tr>
<td>order</td>
<td>0.0087 1 0.92588</td>
<td>26.0099 1 3.397e-07 ***</td>
</tr>
<tr>
<td>predictability: order</td>
<td>0.0780 1 0.77999</td>
<td>4.0696 1 0.04366 *</td>
</tr>
</tbody>
</table>

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

- When given no instruction to imitate, predictability of the target words is significant.
- When told to imitate, predictability is not significant, however order and a predictability*order interaction are.
• When told to imitate, the subjects’ VOT becomes more like the models’ over time for unpredictable words, while there is no change for predictable words.
• Mixed effects model shows significant difference by predictability in the 4th quarter (tokens 75-100), $p = 0.0059$ when told to imitate
Results: Pitch

• Compared pitch of the target syllable (first syllable of [k] initial word) to that of model

• relpitchdiff = (target.pitch_{mod} / utterance.pitch_{mod})– (target.pitch_{subj} / utterance.pitch_{subj})

Given significant differences for men and women, these results are considered separately

Analysis of Deviance Table (Type II Wald chisquare tests)

Response: RELPITCHDIFF ~ PREDICTABILITY * GROUP * ORDER * GENDER + (1|SUBJECT) + (1|WORD)

<table>
<thead>
<tr>
<th></th>
<th>NO INSTRUCTION TO IMITATE</th>
<th>TOLD TO IMITATE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Chisq</td>
<td>Df</td>
</tr>
<tr>
<td>gender</td>
<td>6.0356</td>
<td>1</td>
</tr>
<tr>
<td>predictability:order:gender</td>
<td>0.0725</td>
<td>1</td>
</tr>
</tbody>
</table>

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Pitch: Men

• Similar pattern to VOT: closer to the model for unpredictable words

<table>
<thead>
<tr>
<th></th>
<th>NO INSTRUCTION TO IMITATE</th>
<th>TOLD TO IMITATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictable</td>
<td>-23.82%</td>
<td>-9.06%</td>
</tr>
<tr>
<td>Unpredictable</td>
<td>-17.00%</td>
<td>-7.18%</td>
</tr>
<tr>
<td>RELPITCHDIFF (MEN)</td>
<td>mean</td>
<td>median</td>
</tr>
<tr>
<td></td>
<td>-23.82%</td>
<td>-19.83%</td>
</tr>
<tr>
<td></td>
<td>-17.00%</td>
<td>-15.83%</td>
</tr>
<tr>
<td></td>
<td>-9.06%</td>
<td>-6.37%</td>
</tr>
<tr>
<td></td>
<td>-7.18%</td>
<td>-4.64%</td>
</tr>
</tbody>
</table>
Similar pattern to VOT (all subjects), for Predictability and Predictability:order--- when told to imitate male subjects get closer to the pitch contour but only for unpredictable words (correlations: pred, r = -0.06, unpred, r = 0.2)

Unexpectedly, Order is significant when no instruction is given, and there is a negative correlation--- subjects drift further from the model over time (r = -0.1)
Pitch: Women

- Womens’ pitch on target syllables was higher for unpredictable words--- but actually exceeding the model in relative pitch.

<table>
<thead>
<tr>
<th></th>
<th>NO INSTRUCTION TO IMITATE</th>
<th>TOLD TO IMITATE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predictable words</td>
<td>Unpredictable words</td>
</tr>
<tr>
<td><strong>RELPitchDiff (WOMEN)</strong></td>
<td>mean</td>
<td>-3.55%</td>
</tr>
<tr>
<td></td>
<td>median</td>
<td>-5.01%</td>
</tr>
</tbody>
</table>
The pattern for women imitating pitch was different from men;
Predictability was significant in both experiments
No predictability:order interaction in either experiment
When told to imitate their pitch became more like the model over time
Discussion

• The findings from these experiments show that context modulates attention to phonetic details, where fewer phonetic details are noticed for predictable words.

• Higher level linguistic knowledge therefore can shape our exemplars, agreeing with Goldinger’s (2007) quote:
  • “[E]ach stored exemplar is actually a product of perceptual input combined with prior knowledge…”
The results also suggest something of a ‘how’ and a ‘what’ listening mode, where predictable speech is processed in the ‘what’ mode and unpredictable speech is processed in the ‘how’ mode.

More complicated perhaps? Two extreme endpoints of a scale?
• Both predictability and instructional conditions may push speakers towards different listening mode

• Weight of abstracted form vs. new exemplar in both perception and production

• Predictability*order interaction: perhaps over time subjects became accustomed to prominence pattern, and abstracted these details as well for predictable words
Implications for Sound Change

• The imitation study showed that the difference in perception of words according to predictability affected production as well; Thus relevant for sound change.

• Another dimension of lexical diffusion? How to quantify a word’s global predictability? (as opposed to frequency)

• Perhaps this process drives the profound phonological differences in reliably more predictable word classes ---- such as function words, as opposed to content words

• Could influence or facilitate the development of phrasal prominence? (as opposed to Lindblom’s hyperarticulation /production based account)
Bibliography


