Today

- Top-down and bottom-up influences on speech perception.
- Role of linguistic knowledge (grammar) in speech perception.
Phonemic restoration effect
Perceptual restoration / phonemic restoration effect (Warren, 1970):

When a speech sound is masked by noise (e.g. a cough), listeners ‘restore’ the missing phoneme, often without noticing there is a missing phoneme at all.

Listeners also are generally poor at determining where the noise occurred in the signal.
Bottom-up processing:

Perception starts at the sensory input and moves ‘up’ through successively higher levels of representation. Data-driven, the information flows from ‘bottom’ to ‘top’.

See also: Feedforward
Top-down processing

The interpretation of the sensory input is guided by contextual information given by knowledge about higher levels of representation. Expectation or ‘hypothesis’ driven, the information flows from ‘top’ to ‘bottom’.

See also: Feedback
Bottom-up and top-down

knowledge

/ɪgɪslɛʃɪn/

perception

[lɪgɪslɛʃɪn]

sensory input

[lɪgɪslɛʃɪn]
Analysis by synthesis

→ Analysis by synthesis (Halle & Stevens): Perceivers use a ‘forward model’ that generates the acoustic consequences of higher order representations. Internal simulation is matched against input to determine perception.

→ Analysis by synthesis proposes that listeners hypothesize and test, and that perception reflects those hypotheses that receive sufficient evidence in input.
Analysis by synthesis

→ Proceeds by doing preliminary analysis of input, which then is used to generate candidate hypotheses for input.

→ The acoustic consequences of candidate hypotheses are internally simulated / generated and compared against input.
→ **Perception is (partially) constructive:** We use top-down knowledge to construct our perception of the input.
Bayes’ Rule

**Bayes’ rule:** Rule that relates one conditional probability to its inverse.

\[ P(X \mid Y) \propto P(Y \mid X)P(X) \]
Bayes’ rule: Can take on special importance in modeling perception when we think of it as expressing the probability of some hypothesis (perception) given some data (input). Useful for describing reasoning under uncertainty:

\[
P(\text{Hypothesis} \mid \text{Data}) \propto P(\text{Data} \mid \text{Hypothesis}) \cdot P(\text{Hypothesis})
\]

- likelihood of data given hypothesis
- prior probability of hypothesis
Multiple hypotheses about the input are maintained and ranked/scored in view of the input/data.
Prediction: We should see that top-down linguistic knowledge is used to shape perception in predictable ways across languages.
Analysis-by-synthesis posits a role for active prediction of incoming sensory input, driven by rules that generate structure. Is there any evidence for online prediction of this sort?
Many varieties of English show anticipatory nasalization of vowels preceding nasal consonants:

[aba] [ãma]
?[ama] *[ãba]

Distinction between oral and nasal vowels is not phonologically contrastive in English and does not intuitively seem salient to speakers.

This is a feature of English. It happens in other languages, e.g. Bengali, even though Bengali has contrastive vowel nasalization.

Do English speakers anticipate a nasal consonant immediately following a nasal vowel?
→ **Gating task**: Partial stimulus is presented and participants asked to complete it:

[pã]..

→ **English** speakers regularly predicted nasal consonants following these stems: [pãŋ]..

→ **Bengali** speakers did not! Instead they interpreted nasal vowels as underlying nasals: [pãt]...

→ English speakers anticipate there is only one possible source for nasal vowels - following nasal consonant - and predict it. Bengali speakers do not.
Fig. 2. Example of stimulus spectrograms show congruent [aba] (top) and anomalous [ama] (bottom).
Flagg et al. (2005)

Fig. 2. Example of stimulus spectrograms show congruent [aba] (top) and anomalous [ama] (bottom).
→ Japanese phonotactic restrictions (mostly) only allow the following syllables underlyingly:

V - vowel

CV - consonant + vowel

CVN/Q - consonant + vowel + nasal or geminate consonant

**English:** [mɪk.'dæn.ˈldz]

**Japanese:** [ma.ku.do.na.ru.do]
What happens when input conflicts with grammatical expectations?
→ Compared Japanese speakers and French speakers.

→ Participants asked to identify if /u/ was present in stimulus or not.

→ Japanese speakers reported /u/ was present even when it was not physically present in the stimulus.

→ **Phonological illusion**: A vowel is heard (epenthesized) where it is necessary to break up an illegal consonant sequence.

![Graph showing percentage of [u] vowel judgments in stimuli as a function of vowel duration](image)

*Figure 1.* Percentage (y axis) of [u] vowel judgments in stimuli such as *ebuzo* in French and Japanese participants as a function of vowel duration (x axis) in Experiment 1.
→ Compared Japanese speakers and French speakers.

→ Participants asked to identify if /u/ was present in stimulus or not.

→ Japanese speakers reported /u/ was present even when it was not physically present in the stimulus.

→ **Phonological illusion**: A vowel is heard (epenthesized) where it is necessary to break up an illegal consonant sequence.

*Figure 2*. Percentage (y axis) of [u] vowel judgments in stimuli such as *ebu*zo in French and Japanese participants as a function of vowel duration (x axis) in Experiment 2.
→ **Perceptual assimilation:** non-native sounds (or groups of sounds) may be perceptually ‘assimilated’ to the nearest phonologically legal category, through acoustic or articulatory similarity (Best, 1995).

→ Japanese has a process of high vowel devoicing between unvoiced consonants:

→ /i/ and /u/ are arguably ‘perceptually closest’ to a very reduced (or missing) vowel.

/ga.ku.sei/ ‘student’  
/a.ki.ta/ ‘bored’  

[ga.k(ʊ).sei]  
[a.k(ɨ).ta]
English: 
[mɪk.'dæn.ʃdz]

Brazilian Portuguese: 
[ma.ki.do.'nau.dʒiʃ]

Japanese: 
[ma.ku.do.na.ru.do]
In BP, /i/ is the shortest vowel; in Japanese /u/ is.

Both BP and Japanese speakers show illusory vowel epenthesis effects...

Fig. 2. Mean percentages ‘vowel present’ responses as a function of cluster type in Japanese, Brazilian Portuguese, and European Portuguese participants (Experiment 1). Error bars represent standard errors.
Brazilian Portuguese (BP) has the same high vowel devoicing process as Japanese.

In BP, /i/ is the shortest vowel; in Japanese /u/ is.

... but Brazilian speakers generally reported /i/, whereas Japanese speakers generally reported /u/.

Fig. 2. Mean percentages ‘vowel present’ responses as a function of cluster type in Japanese, Brazilian Portuguese, and European Portuguese participants (Experiment 1). Error bars represent standard errors.
→ Critical stimuli can be made by splicing out /i/ or splicing out /u/. Does this make a difference?

→ Even after splicing out /u/ or /i/, the coarticulatory influence of the original vowel identity may remain in the surrounding segments.

→ This contributes weak bottom-up cues to vowel identity.
Two stage repair (e.g. Berent et al., 2007): Input is processed bottom-up and eventually reaches grammar, where it is repaired:

[ebzo]u → /ebzo/ → /ebuzo/

Phonological analysis: Recode into phonemes.

Phonological repair: Use phonotactic constraints to repair.
Single-stage vs. two-stage repair

→ **Two stage repair** (e.g. Berent et al., 2007): Input is processed bottom-up and eventually reaches grammar, where it is repaired:

\[
 [\text{ebzo}]_u \xrightarrow{\text{Phonological analysis: Recode into phonemes.}} /\text{ebzo}/ \xrightarrow{\text{Phonological repair: Use phonotactic constraints to repair}} /\text{ebuzo}/
\]

→ **Single-stage**: Phonological analysis is jointly influenced is processed bottom-up and eventually reaches grammar, where it is repaired:

\[
 [\text{ebzo}]_u \xrightarrow{\text{Phonological interpretation: Balance top-down (grammatical) and bottom-up (acoustic) constraints.}} /\text{ebuzo}/ \xrightarrow{\text{}} /\text{ebuzo}/
\]
Interplay of top-down and bottom-up

... Brazilian and Japanese contrast holds for all stimuli: always more /i/ responses for Brazilians than Japanese.
Interplay of top-down and bottom-up

... but residual phonetic cues systematically influence rate of /i/ or /u/ for both languages.
→ **Single-stage**: Phonological analysis is jointly influenced and eventually reaches grammar, where it is repaired:

\[ [\text{ebzo}]_u \rightarrow /\text{ebuzo}/ \leftarrow /\text{ebuzo}/ \]

**Phonological interpretation:**
Balance top-down (grammatical) and bottom-up (acoustic) constraints.

→ **What kind of top-down constraints can apply?**

- Phonotactic constraints (legal syllable structures)
- Language-wide generalizations about ‘minimal vowels’
- **Phonological rules / alternations?**
Korean /ɨ/ is shortest and participates in vowel deletion processes: Good candidate for illusory vowel

(1) Relevant phonological processes in Korean

a. Vowel deletion

| /i/ → Ø / __ + V | /khɨ+eto/ → [khədo]^{10} | ‘although (it is) big’ |

or

| /i/ → Ø / V + __ | /khɨ+iňi/ → [khani] | ‘because we go’ |

b. Palatalisation

| C[alv] → [pal] / __ i | /phɨ+i/ → [pacho] | ‘dry field (NOM)’ |
| | /os+i/ → [ofi] | ‘clothes (NOM)’ |
Korean /ɨ/ is shortest and participates in vowel deletion processes: Good candidate for illusory vowel

(1) Relevant phonological processes in Korean

a. Vowel deletion

\[
/ɨ/ \rightarrow \emptyset / \_ + V \\
/kʰɨ+əto/ \rightarrow [kʰədo]^{10} \quad \text{‘although (it is) big’}
\]

or
\[
/ɨ/ \rightarrow \emptyset / V + \_ \\
/kʰa+ini/ \rightarrow [kʰani] \quad \text{‘because we go’}
\]

b. Palatalisation

\[
[C] \rightarrow [pal] / \_ i \\
/pa^h+i/ \rightarrow [pa^chi] \quad \text{‘dry field (NOM)’}
\]
\[
/os+i/ \rightarrow [ofi] \quad \text{‘clothes (NOM)’}
\]

/i/ palatalizes preceding consonants. Importantly, /ʃ/ can only occur before /ɨ/.

Durvasula & Kahng (2015)
Grammatical knowledge: Alternations

Vowel perceived by Korean speakers reflects inference based on phonological alternations.

Durvasula & Kahng (2015)
Finer-grained knowledge: Statistical cues

→ **Statistical knowledge**: Do we encode record fine-grained information about the likelihood of different linguistic events? Is this part of our linguistic knowledge?

→ **(Forward) transitional probability**: Probability of one item following another, written as $P(B | A)$ (probability of $B$ given $A$).

→ We can use transitional probability to quantify the likelihood that a given phoneme follows another in a language:

<table>
<thead>
<tr>
<th></th>
<th>/g/</th>
<th>/ʃ/</th>
<th>/tʃ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>19,439</td>
<td>1,740</td>
<td>478</td>
</tr>
<tr>
<td>/i/</td>
<td>1,168</td>
<td>26,530</td>
<td>7,219</td>
</tr>
<tr>
<td>/u/</td>
<td>2,087</td>
<td>1,033</td>
<td>23</td>
</tr>
<tr>
<td>/e/</td>
<td>2,258</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>/o/</td>
<td>5,441</td>
<td>1,715</td>
<td>995</td>
</tr>
<tr>
<td>Total</td>
<td>30,393</td>
<td>31,032</td>
<td>8,730</td>
</tr>
</tbody>
</table>

*Table 2. Relevant co-occurrence counts from the CSJ-RDB.*

Kirkland et al (2020)
→ **Statistical knowledge**: Do we encode record fine-grained information about the likelihood of different linguistic events? Is this part of our linguistic knowledge?

→ **(Forward) transitional probability**: Probability of one item following another, written as $P(B|A)$ (probability of $B$ given $A$).

→ We can use transitional probability to quantify the likelihood that a given phoneme follows another in a language:

<table>
<thead>
<tr>
<th></th>
<th>/g/</th>
<th>/ʃ/</th>
<th>/tʃ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>19,439</td>
<td>1,740</td>
<td>478</td>
</tr>
<tr>
<td>/i/</td>
<td>1,168</td>
<td>26,530</td>
<td>7,219</td>
</tr>
<tr>
<td>/u/</td>
<td>2,087</td>
<td>1,033</td>
<td>23</td>
</tr>
<tr>
<td>/e/</td>
<td>2,258</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>/o/</td>
<td>5,441</td>
<td>1,715</td>
<td>995</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>30,393</td>
<td>31,032</td>
<td>8,730</td>
</tr>
</tbody>
</table>

→ **For example**: $Pr(/a/|/g/)$ in Japanese is .64, $Pr(/a/|/S/) is 0.6$ according to Kirkland et al (2020).
Finer-grained knowledge: Statistical cues

[etʃpo]  etʃipo (46%)
[eiʃpo]  eiʃipo (31%)
[egpo]   egipo (0.3%)

**Table 2.** Relevant co-occurrence counts from the CSJ-RDB.

<table>
<thead>
<tr>
<th></th>
<th>/g/</th>
<th>/ʃ/</th>
<th>/tʃ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>19,439</td>
<td>1,740</td>
<td>478</td>
</tr>
<tr>
<td>/i/</td>
<td>1,168</td>
<td>26,530</td>
<td>7,219</td>
</tr>
<tr>
<td>/u/</td>
<td>2,087</td>
<td>1,033</td>
<td>23</td>
</tr>
<tr>
<td>/e/</td>
<td>2,258</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>/o/</td>
<td>5,441</td>
<td>1,715</td>
<td>995</td>
</tr>
<tr>
<td>Total</td>
<td>30,393</td>
<td>31,032</td>
<td>8,730</td>
</tr>
</tbody>
</table>

Kirkland et al (2020)
Finer-grained knowledge: Statistical cues

$P(i | \text{ṣ}) = 0.85)$
$P(i | \text{ʃ}) = 0.82)$

etʃipo (46%)
evʃipo (31%)
eggi (0.3%)

Table 2. Relevant co-occurrence counts from the CSJ-RDB.

<table>
<thead>
<tr>
<th></th>
<th>/g/</th>
<th></th>
<th>/ʃ/</th>
<th></th>
<th>/tʃ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>19,439</td>
<td>1,740</td>
<td>478</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/i/</td>
<td>1,168</td>
<td>26,530</td>
<td>7,219</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/u/</td>
<td>2,087</td>
<td>1,033</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/e/</td>
<td>2,258</td>
<td>14</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/o/</td>
<td>5,441</td>
<td>1,715</td>
<td>995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>30,393</td>
<td>31,032</td>
<td>8,730</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
→ **Perceptual assimilation**: non-native sounds (or groups of sounds) may be perceptually ‘assimilated’ to the nearest phonologically legal category, through acoustic or articulatory similarity (Best, 1995).

→ **Syllable initial /tl/:** Licit in many languages (e.g. Hebrew, Mexican Spanish), illicit in others (e.g. English, French).

![Diagram showing the pronunciation of /tl/ in Hebrew and English](hebrew-vs-english.png)
Perceptual assimilation

Halle & Best (2007)
Perception involves integration of top-down and bottom-up information.

Top-down influence means linguistic perception is partially constructive: Our internal models of our language ‘fill in the blanks’ to help interpret the input.

Phonological illusions driven by grammatical and statistical knowledge are one example of this.
**ABX Discrimination procedure:** Play two stimuli in a row (A + B), and then present a third. Participant is asked to identify if X is the same as A, or B.

A:  
X:  
B:
ABX Discrimination procedure: Play two stimuli in a row (A + B), and then present a third. Participant is asked to identify if X is the same as A, or B.

A:

X:

B:
→ 32% error rate for Japanese speakers on *ebuzo - ebzo* contrast!
(Compare to 50% chance rate - pure guessing).

→ Longer reaction times to make decision - less confident / certain decisions.

→ French speakers had low error rates, and faster reaction times (RTs).

→ The opposite pattern was seen for a phonotactic contrast that Japanese has, but French does not (vowel length).

---

**Table 1**

*Mean Reaction Time, Standard Error, and Error Rate in ABX Judgments on an Epenthesis Contrast and a Vowel Length Contrast in French and Japanese Participants in Experiment 3*

<table>
<thead>
<tr>
<th>Language</th>
<th>Vowel length contrast <em>(ebu-zo–e-bu-uzo)</em></th>
<th>Epenthesis contrast <em>(e-bu-zo–e-bzo)</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT (ms)</td>
<td>SE</td>
</tr>
<tr>
<td>Japanese</td>
<td>1,082</td>
<td>45</td>
</tr>
<tr>
<td>French</td>
<td>1,173</td>
<td>73</td>
</tr>
</tbody>
</table>

*Note.* RT = reaction time.