Phonetic Accommodation: Who is speaking and are we really listening?

Molly Babel

University of British Columbia

Sociolinguistic Variation and Language Processing
March 31, 2016
Acknowledgements

- Research participants: thank you!
- Funders (SSHRC, CFI, UBC, UCB, NSF): thank you!
- Collaborator: Grant McGuire (UC Santa Cruz)
We learn how to listen.

- We learn the perceptual categories in our native languages (e.g., Werker & Tees, 1984)
- This is finely tuned: infants’ perceptual contrasts related to degree of caretakers’ contrasts in production (Cristià, 2011).
- As adults, this means listeners with different language backgrounds attend to and prioritize different parts of the phonetic signal (e.g., Francis & Nusbaum, 2002).
- Not all input has the same influence on all listeners.
- Different styles of listening – perception- or comprehension-oriented attentional sets.
  - Attentinal sets give you different levels of detail from the speech signal (e.g., Culter et al., 1987; McAuliffe, 2015)
Selective patterns in children

- Acquisition of gender-specific speech patterns before major anatomical differences surface (Sachs et al. 1973).
- Children tend to adopt local dialect patterns (e.g., Trudgill, 1981)
Kids seem to either learn selectively or show biases in selecting production variants.

It doesn’t seem like primary exposure or quantity of exposure is the always the determining factor.
Adults shift when exposed to new dialects, but they rarely fully adapt new speech patterns (Munro et al., 1999).

**Why?** Exposure to novel information (and *that grabs attention, triggers change*) or is it positive social evaluation?
We learn our linguistic contrasts, and we learn cultural and social meaning.
Hearing a man’s voice say “I might be pregnant” generates a neurological response that’s indicative of surprisal (Van Berkum et al., 2008).

Hearing natively accented Dutch say an ungrammatical utterance generates an “ungrammatical” neurological response, but hearing the same by a non-native speaker does not (Hanulíková et al., 2012).
It’s not all cool.

- Stereotypes about southern accents in adults and children (Preston, 1993; Kinzler & DeJesus, 2013).

Kinzler & DeJesus (2013)
We do linguistic analysis.
We do social evaluation of voices.
Does that social evaluation affect what we get out of a voice linguistically?
Accommodation paradigms let us look at some of these issues.
Convergence is the act of one’s speech patterns becoming more similar to speech one has been recently exposed to.

- It signals some sort of relationship between speech perception and production.
- In phonetics, it suggests sensitivity to phonetic details.
- What causes/allows/facilitates it?
What’s behind it?

- Is it to facilitate social interaction and decrease social distance? (e.g., Giles, 1973)
- Is it about the organization of representations across speech perception and production? (e.g., Goldinger, 1998; Sancier & Fowler, 1997)
Probably both.

- Is it to facilitate social interaction and decrease social distance? (e.g., Giles, 1973)
- Is it about the organization of representations across speech perception and production? (e.g., Goldinger, 1998; Sancier & Fowlder, 1997)

- Cognitive and social factors **both** seem to play an important role (e.g., Pardo, 2006).
Listener repertoire and attitudes

**Repertoire:** If you are merged, you don’t decrease your mergedness (Babel et al., 2013).
Attitudes: Positive attitudes toward the nationality of the speaker/accent correlates with an increase in accommodation (Babel et al., 2013).
Voices and listeners

- How about different voices? How do they elicit different kinds of behavior?
- Namy et al. (2002) that one particular voice elicited more accommodation than others.
- Considering voices used in these tasks let’s us examine how we might listen differently to different kinds of voices.
Voices: 60 voices (Babel, 2009, 2012)
Vocal aesthetics: 30 listeners (Babel, McGuire, & King, 2014; Babel & McGuire, 2015).
- Perceived Attractiveness on Likert Scale
- Gender Categorization Fluency
Voice Selection
Predictions

- More Convergence
  - Less Typical voices: novelty preferences
  - More Attractive voices: social preferences

- Less Convergence
  - More Typical: nothing out of the ordinary
  - Less Attractive: socially disfavored
**Shadowers** (10 male, 10 female) completed an auditory naming task.
- Baseline productions and single-word shadowing in response to the 8 model talkers.

**Listeners** (n=159) quantified convergence in an AXB task.
- Baseline and shadowed tokens are compared as AB tokens to a model’s X.
Not all voices are responded to equivalently.

<table>
<thead>
<tr>
<th>Model Gender x Voice Type x Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attractive</td>
</tr>
<tr>
<td>Typical</td>
</tr>
<tr>
<td>Shadowed tokens judged as more similar</td>
</tr>
<tr>
<td>More Convergence</td>
</tr>
<tr>
<td>Less Convergence</td>
</tr>
<tr>
<td>More Convergence</td>
</tr>
<tr>
<td>Less Convergence</td>
</tr>
</tbody>
</table>

ModelGender
- Female
- Male
Female and male shadowers respond to the voices differently.
Summarizing the figures

- Females imitated more, but that depended on voice type.
- Male and female shadowers accommodated the least typical voices equally.
- Males did not accommodate to any voices that were selected on the attractiveness dimension.
- Females accommodated to the most attractive voices more than the least attractive voices.
Females generally show a greater ingroup bias (Rudman and Goodwin 2004), which might contribute to females showing more social preferences in accommodation.

Namy et al. (2002) suggest that socialized practices may lead females to attend more to indexical features, which they argue leads to more accommodation by females.

Females may be more sensitive to vocal prestige (e.g., Labov, 2001, but see Schilling-Estes, 2002 for criticisms).
Some listeners may pay more attention to the details of speech.

Some speakers may elicit more attention than others.

Speech behavior is an interplay between the listener and the talker.
Social biases or increased sensitivities could lead to listeners encoding select speech in more detail, facilitating accommodation.

The idea of an attention weighting mechanism in exemplar models is (relatively) old (Johnson 1997).
Own-race bias in face memory:
- Encoding the right cues and doing so efficiently (Meissner & Brigham, 2001)
- Minimal groups designs show that it is also about attention and effort (Young et al., 2012; Van Bavel & Cunningham, 2012)
Emotionally positive and emotionally negative images are experienced more vividly (Todd et al., 2012).
Some visual experiences are enhanced relative to others.

Figure 1. a. NE task to assess perceptual vividness. A standard, overlaid with a constant level of noise, was given an arbitrary value of 100. Each standard was followed by a positive, negative, or neutral image overlaid with one of 3 levels of noise. In each trial, participants entered a number representing an estimate of the proportion of noise on the image relative to the standard. b. Emotionally salient images were psychophysically scaled to contain less noise, i.e., were perceived as more perceptually vivid, despite equal levels of objective noise. The right side of the image illustrates a 15% decrement in noise level from the left. c. NE ratings for positive, negative and neutral pictures at varying levels of objective noise: 1 is less than standard, 2 is the same as standard, and 3 is greater than standard.
Theories behind visual processing illustrate the flexibility with how we perceive and process the visual world.

There are also different ways of listening to speech.

Listeners have some control over perceptual processes (e.g., talker normalization: Magnuson & Nusbaum, 2007 Barreda, 2012).
Linguistic and social signals are intertwined in the same medium.

We need to think about the voices we use in speech experiments, as listeners may attend and process different voices differently.
Thanks for listening!