Perceptual Adjustments: All Pronunciations are not Created Equal

Tanya Kraljic, Susan E. Brennan, Arthur G. Samuel

State University of New York at STONY BROOK Department of Psychology

UCSD

Acknowledgements

This material is based upon work supported by a National Science Foundation Graduate Fellowship and by NSF Grant No. 0325188 and NIMH RO151663.

Introduction

People pronounce sounds differently for different reasons. Two ubiquitous reasons are:

1. The Context (e.g., Liberman et al., 1954) Contextually-driven changes are pervasive in dialects.

Examples of (contextually-driven) dialectal variations

R-dropping (Where did you pahk the car?): Boston speakers drop /r/'s after vowels, but never when /r/ is the first sound in a word, or when it occurs after a consonant

T-flapping (latter [ladr] = ladder [ladr]): American English speakers 'flap' /t/'s when they occur intervocalically, but never when they start a stressed syllable

Place assimilation (/s/ = ?s∫ before a [tr] cluster, as in 'street' or 'construction'): In many Northeastern US dialects, /s/ becomes like /∫/ (sh), but only before a [tr] cluster

2. The Speaker (e.g., Peterson & Barney, 1952) – A speaker's age, gender, and linguistic experience will determine the acoustic realization of any particular sound.

Examples of (speaker-driven) idiolectal variations

Lisps, *foreign accents*, *temporary physical states* (food in one's mouth, being drunk)

The Question: How do listeners handle variation in pronunciation?

Previous Research: Psycholinguistics

1. Listeners get rid of variability: Evidence based on contextually-driven variation

How might this work? One option: feature parsing (Gow et al.):
Listener hears a sound that's assimilated: [?s∫] in *[street]*

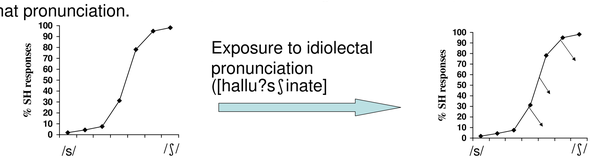
They decode the features present in that sound: Some features are consistent with /s/, other (weaker) features are consistent with /∫/

So they 'assign' the weaker features to the subsequent sound:

[?s∫] → (weaker features) → [tr] = /s/ + [tr]

2. Listeners represent variability: Evidence based on speaker-driven variation

How might this work? One option: perceptual learning:
'odd' pronunciations result in phonemic categories that are expanded to accommodate that pronunciation.



Previous Research: Sociolinguistics

The different psycholinguistic accounts are based on different types of variation, but speech recognition models do not consider the source of a variation. Perhaps these models should? After all, listeners appear to:

3. Listeners use variability to form attributions about speakers: Speech perception, in turn, may be mediated by listeners' beliefs about the speaker

- Listeners' perceptual boundaries for a fricative contrast (/s/-/∫/) or a vowel contrast (/U/-/ʌ/) shifted depending on whether they saw a male or a female face 'producing' syllables which contained those sounds (Strand and colleagues)
- Listeners perceived vowels differently depending on what they believed to be the dialect of the speaker they were listening to (Niedzielski, 1999)

Our Research

Question

Does the source of a particular variation determine how the perceptual system adjusts to it?

Specifically: Is perceptual learning the same after exposure to **Dialectal** variation as after exposure to **Idiolectal** variation?

?s∫ as an Idiolect vs. as a Dialect – Background

Idiolectal variation (not contextually-constrained) → perceptual learning:

- /s/ in words like *hallucinate*, *obscene*, *parasite*
- /s/ replaced with a sound midway between /s/ and /∫/ (?s∫)
- Listeners expand their category for /s/ to accommodate this idiolectal pronunciation (Kraljic & Samuel, 2005)

Dialectal variation (contextually-constrained) → ?? → perceptual learning?

- NY/LI dialect: /s/ becomes ?s∫ when it immediately precedes [tr] (*street*)
- Acoustically: Idiolect ?s∫ and Dialect ?s∫ are identical
- Perceptual system's task is also identical: Recognize the underlying /s/
- ONLY the source of the variation differs (*Idiolectal vs. Dialectal*)

Experiment Method

Experimental Groups

Phase 1: Exposure (Lexical Decision) – Two groups

Purpose: To expose listeners to a speaker whose pronunciation of /s/ is ambiguous (?s∫)

Manipulation: Ambiguity was attributable either to an Idiolect or to a Dialect

1. *Idiolectal pronunciation*: [?s∫] replaced /s/ in a variety of lexical contexts (e.g., *hallucinate, obscene*)
2. *Dialectal pronunciation*: [?s∫] replaced /s/ only in when it immediately preceded [tr] (e.g., *construction, abstract*)

Phase 2: Category Identification Test

Purpose: to assess whether perceptual learning occurred for each exposure group

Task: All participants categorized items on a VCV continuum; C ranged /s/-like to /∫/-like

All participants: [asi] – [a∫i] continuum

Control Group

Phase 1: None

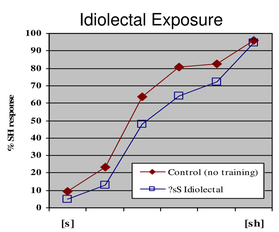
Phase 2: Same as experimental groups

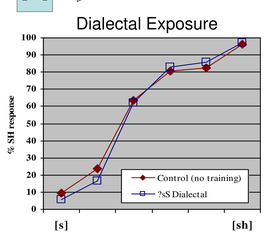
Purpose: To provide a baseline to which experimental groups' responses were compared. If exposure resulted in perceptual learning, participants in the experimental group(s) should have significantly more /s/ responses than those in the control group.

Experiment Results

Listeners exposed to (?s∫) as an Idiolectal variation → perceptual learning

Listeners exposed to (?s∫) in Dialectal contexts → X → no perceptual learning





Why does dialectal exposure not lead to perceptual learning?

We think: Dialectal and Idiolectal variation are processed differently.

But one could argue that the lack of an effect might simply reflect:

- Many of our participants speak this dialect so…no need to adjust [str]?
- Perceptual learning is context-specific and so…the [asi]-{ashi} categorization task may have failed to measure adaptation to dialectal variation?

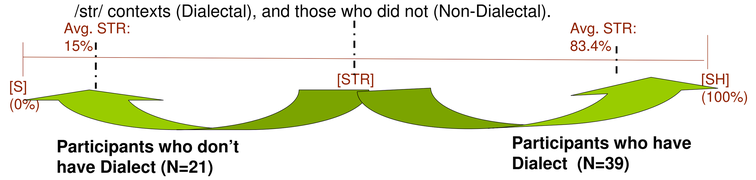
Lack of perceptual learning: Because participants have this dialect? NO

To investigate, we had participants from the experiment read aloud 10 sentences that contained words with /s/, /∫/ and /str/.

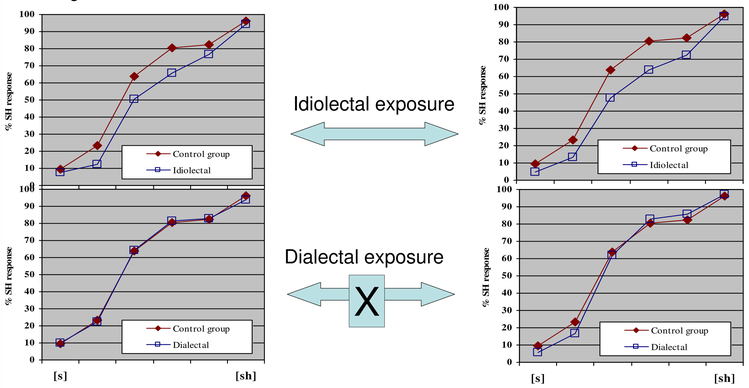
We obtained the mean frequency for each participant's /s/ (S), /∫/ (SH), and the /s/ in /str/ (STR). Within a speaker, /∫/ is always lower frequency than /s/.

We calculated the difference between each participant's S and their STR as a percentage of the difference between their S and SH. A person's STR production could then be assigned a value of 0% (perfectly equal to their S pronunciation), a value of 100% (equal to their SH pronunciation), or any number in between.

On the basis of these frequencies, we split participants into two groups: Those who exhibited ?s∫ in /str/ contexts (Dialectal), and those who did not (Non-Dialectal).



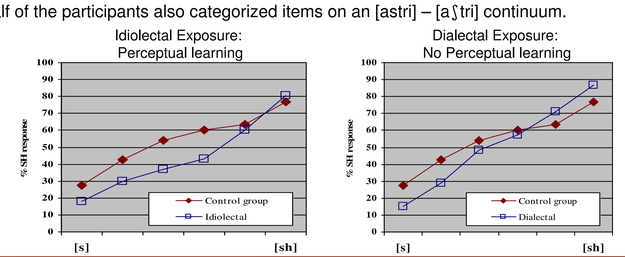
Both groups showed identical patterns of results: Perceptual learning of idiolectal variation, no learning for Dialectal variation:



Participants' own dialect doesn't account for the difference in perceptual learning

Lack of perceptual learning: Dialectal learning not captured by [asi]-[a∫i]?

Half of the participants also categorized items on an [astri] – [a∫tri] continuum. NO



Conclusions

All acoustic-phonetic variations are not treated identically by the perceptual system.

- Exposure to idiolectal pronunciation → general retuning of phonemic categories, even for phonetic contexts that were not in the exposure set.
- Exposure to the same pronunciation in the context of a dialect → X → no perceptual learning.