Perceptual Learning

Given variability (dialect, accent) across speakers, how do we adjust?

If a talker's idiolect has an /ʃ/ that is intermediate between most talkers' /s/ and /ʃ/, what do listeners do?

Example Study

Part 1 – Lexical decision task (Learning)

Part 2 – Phonetic identification task (Testing)

Part 1 – Lexical decision

Listeners hear individual words and nonwords. For each utterance, they decide (word or not) and respond as fast as possible.

Embedded in list of words are 20 critical /s/ words and 20 critical /ʃ/ words. For one group, the /s/ has been replaced with an /s/ + /ʃ/ mixture. Since only /s/ makes these a word, and listeners identify these as words, they are "learning" that this talker makes /s/ a particular way. The 20 critical /ʃ/ words have a clear /ʃ/.

Part 1 – Lexical decision (2)

Second group has the critical words with /ʃ/ modified to have the /s/ + /ʃ/ mixture while the other critical /s/ words have a clear /s/. For the mixed fricative items, only /ʃ/ makes a word and listeners do classify these items correctly. This group is "learning" that the talker produces a somewhat different /ʃ/.

/s/ words	/ʃ/ words
legacy	parachute

Part 2 – Phonetic categorization

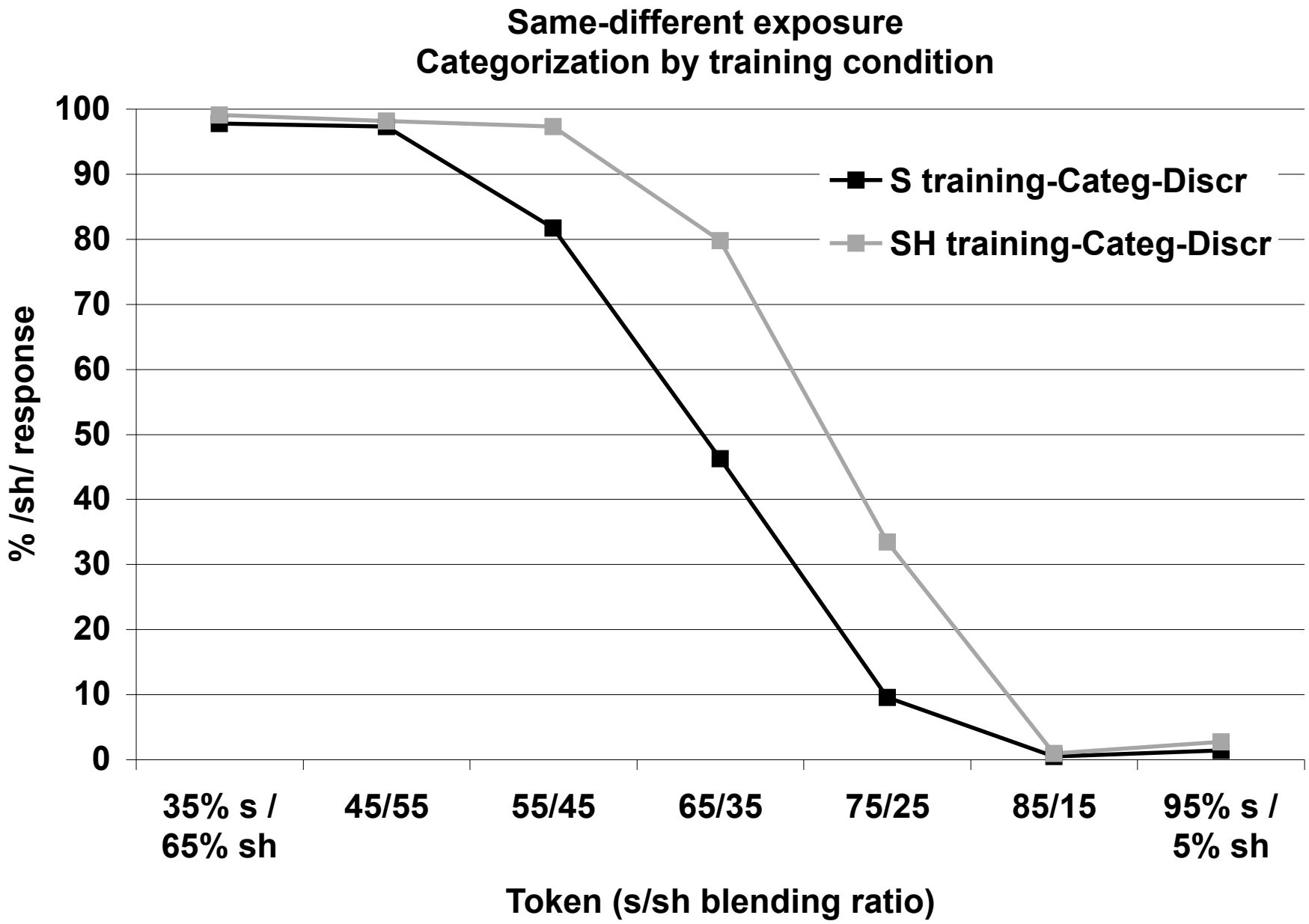
Both groups do same task. They hear syllables that vary from /asi/ to /aʃi/ (fricatives mixed in proportions from entirely or mostly /s/ to entirely or mostly /ʃ/).

For each syllable, they classify the fricative as /s/ or /ʃ/.

Possible Results

If listeners “learn” from the ambiguous fricative that is embedded in a clear, lexical context, it should alter their classification of the /asi/ - /aʃi/ test series.

The group trained that the ambiguous fricative is /s/ should classify more of the test series as /s/ (an enlarged /s/ category) relative to the /ʃ/ trained group.


Same-different exposure
Categorization by training condition
100
90
80
70
60
50
40
30
20
10
0
% /sh/ response
S training-Categ-Discr
SH training-Categ-Discr
35% s / 65% sh
45/55
55/45
65/35
75/25
85/15
95% s / 5% sh
Token (s/sh blending ratio)

Nature of Perceptual Learning

Learning task does not have to be lexical decision. Clark-Davidson et al. used a same-different task and obtained results similar to those of lexical decision. (see previous figure)

The change in the identification category boundary also shows up in changes in AXB discrimination. A signal detection analysis shows that there are underlying changes in sensitivity.

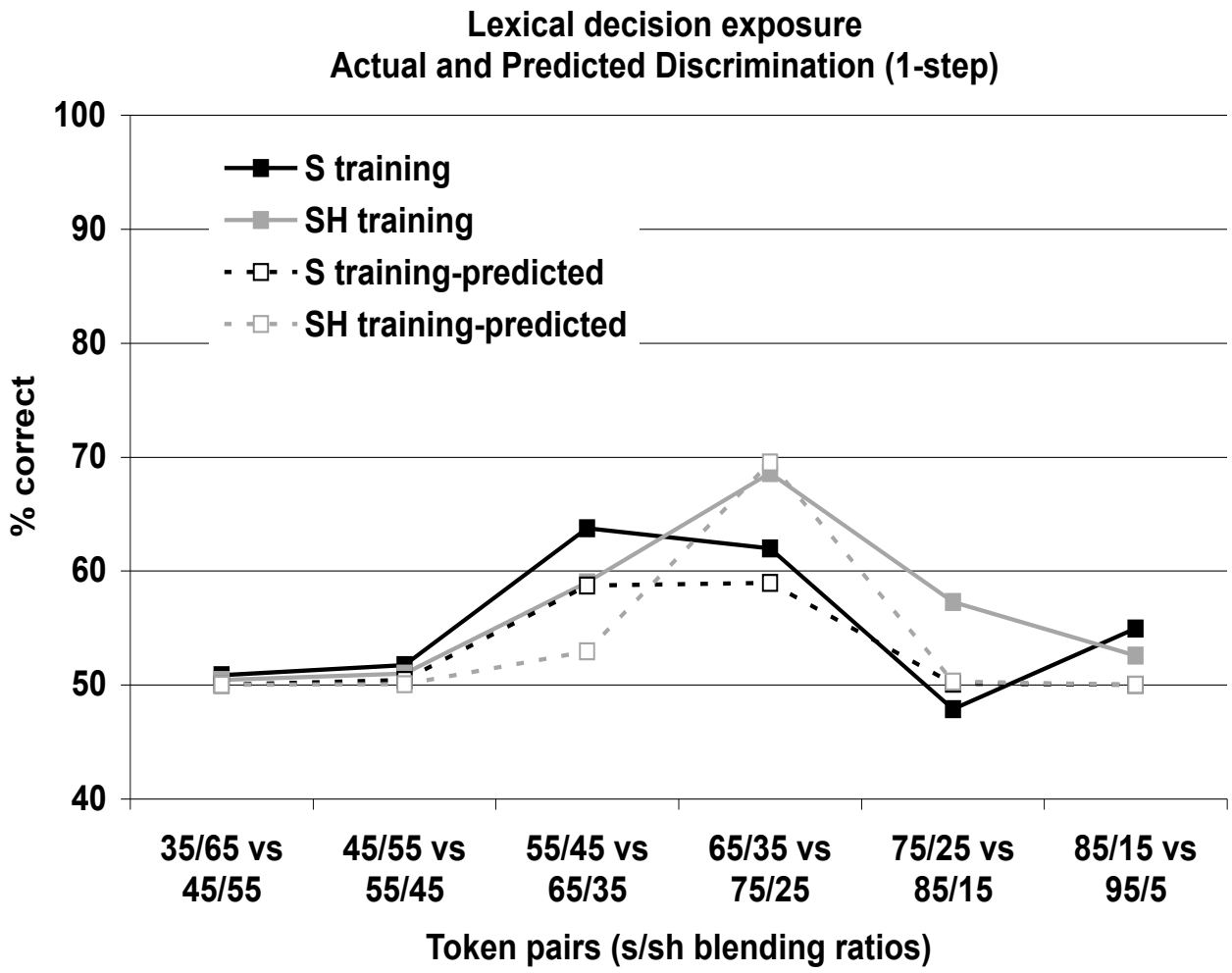
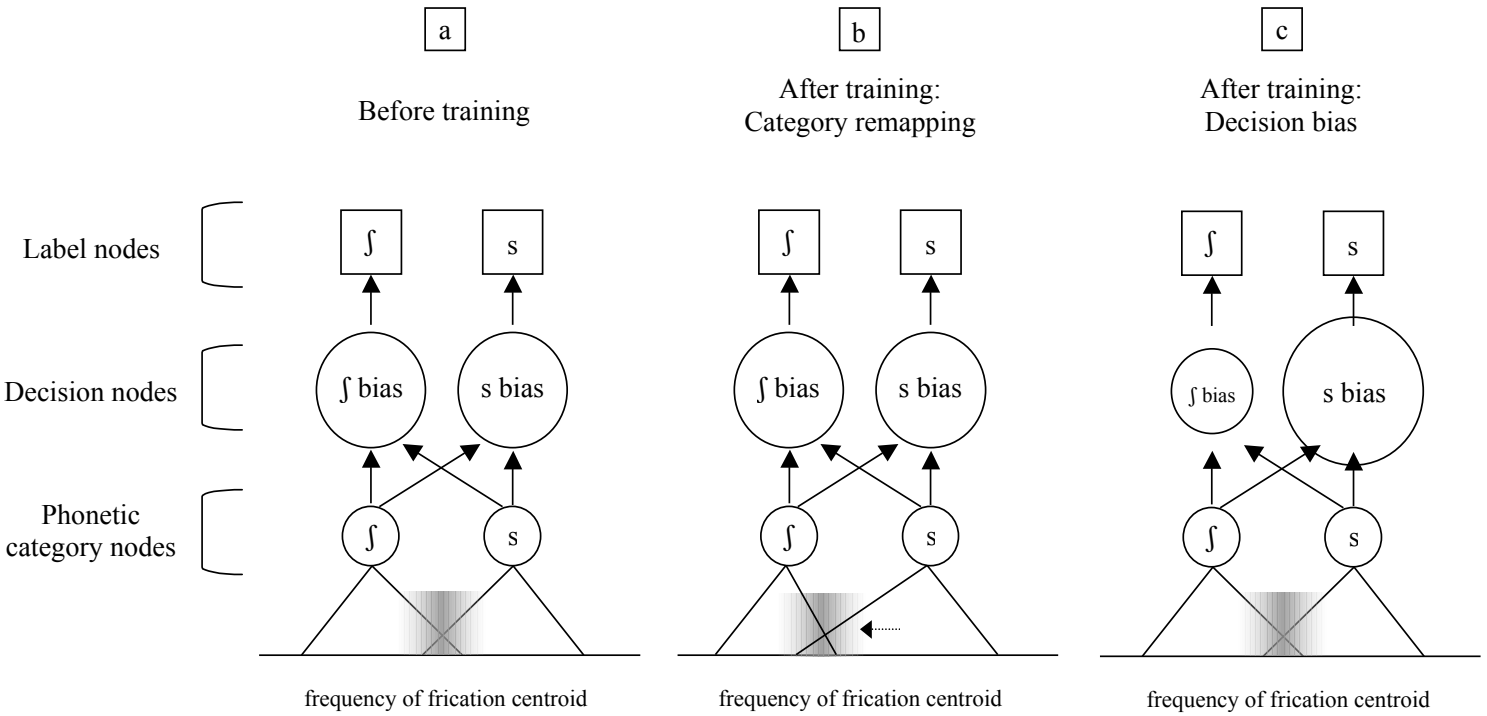

Lexical decision exposure
Actual and Predicted Discrimination (1-step)
100
90
80
70
60
50
40
% correct
S training
SH training
S training-predicted
SH training-predicted
35/65 vs 45/55
45/55 vs 55/45
55/45 vs 65/35
65/35 vs 75/25
75/25 vs 85/15
85/15 vs 95/5
Token pairs (s/sh blending ratios)

Figure 6


a
Before training
b
After training:
Category remapping
c
After training:
Decision bias
Label nodes
Decision nodes
Phonetic
category nodes
ʃ
s
ʃ bias
s bias
frequency of frication centroid
frequency of frication centroid
frequency of frication centroid

Conclusion 1

Listeners are re-tuning or adjusting their perceptual (phonetic) representation. Put another way, effect is not simply a change in “bias”.

Robustness – Talker Specificity

With *fricatives*, the learning is *talker specific*. That is, if trained on one talker, perceptual learning only shows up in testing with the same talker. If trained in opposite directions with two different talkers, can get opposite effects in two different test series.

Robustness – Talker Specificity

With a *VOT contrast in stops*, the learning is *not talker specific*.

If trained with one talker, listeners generalize to another talker.

If trained on a /d/ - /t/ contrast, listeners generalize to a /b/ - /p/ contrast.

Talker Specificity Conclusion

When the contrast represents something that varies with talker (e.g. due to vocal tract length), learning is talker specific.

When contrast might be talker independent (the timing cue of VOT), effect seems to generalize to other talkers.

Robustness - Durability

How fragile or durable is this re-tuning. Put another way, is it easy to produce? Once it occurs, how long does it last? What disrupts it?

Durability - Duration

In one study, after perceptual learning, there was a 12 hour delay prior to testing. For half of listeners, this occurred overnight (sleeping, memory consolidation).

The change in identification was present after a 12 hour delay. It was of equivalent magnitude to immediate test. There was no influence of sleep versus awake in delay interval.

Durability – Duration (2)

If listeners are tested (continuously) to assess how long the effect of perceptual learning lasts, the effect disappears “fairly quickly”. That is, if we were to test them for 60 minutes, the influence would be present in the first 15 minutes, gone by the last 15 minutes.

Repeated exposure to the full range of tokens “re-tunes” the perceptual learning. Put another way, the listener is learning, during the testing, that the talker actually produces the full range.

Durability – Failure to learn

If listeners are pre-tested with the phonetic categorization task then exposed to the lexical decision task, they do not show any perceptual learning.

Apparently, the items from the talker that they are exposed to first have a substantial influence.

Durability – Failure to learn (2)

If listeners have an alternative (to the idiolect of the talker) that the “ambiguous” or “altered” nature of the acoustic-phonetic information can be attributed to (e.g., talker was holding a pen in his/her mouth), then there is no perceptual learning.

If the altered signal can be attributed to a language process (e.g. place assimilation of /flaet maet/ to [flaep maet]), there is no perceptual learning.

Robustness – What can drive the learning process?

If the ambiguous segment is presented in an unambiguous lexical context (e.g. /s/ in “legacy”), this will engender learning. Task can be lexical decision, AX name matching, etc.

If ambiguous segment is presented in a sequence where only one of phonemes is permitted (phonotactic constraint), this will engender learning.

If ambiguous segment is presented with unambiguous visual information (audio-visual integration), this will engender learning.

What next?

Recent study contrasted learning for a phoneme in the middle of a word versus a phoneme at the beginning. No learning for phoneme at beginning. Why?

Are there systematic differences across different phonetic contrasts (e.g. generalization of VOT to other talkers)? Is this related to nature of variability in talkers?

Is this related to other normalization processes?