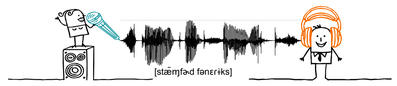
Early effects of speaker gender in spoken word processing

Ed King and Meghan Sumner
Department of Linguistics, Stanford University



How quickly do listeners incorporate information about speaker sex while processing spoken words?

Background

Talker sex contributes to phonetic variation.
Listeners are sensitive to this variation.

Shifting perceptual categories based on talker sex:
- fricative center of gravity (Strand & Johnson, 1996)
- vowel formants (Johnson et al, 1999)

Faster responses when speaker sex matches sex bias:
- sex-biased words in corpora (Hay & Walker, 2013)

Speaker information is thought to be **not incorporated early**.

Speaker effects only emerge in short-term encoding tasks:
- when processing is slower (Luce et al, 2003)
- when processing is difficult (McLennan & Luce, 2005)

Work on the time-course of speaker information has focused on:
- short-term encoding tasks
- individual, speaker-specific, representations

But there is a qualitative difference between
- **short-term**, speaker-specific representations
- **long-term**, group-based representations

When listeners have little information about a group
- reliance on **exemplars**

When listeners have a lot of information about a group
- reliance on **stereotypes**

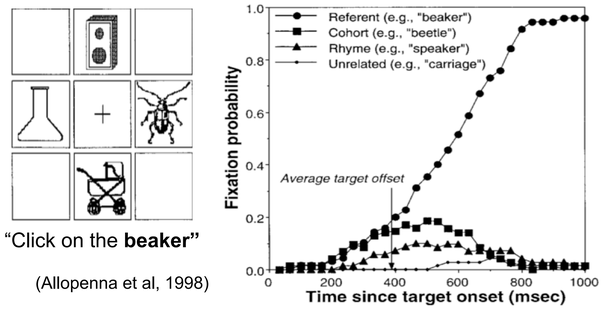
(Sherman, 1996)

We do not know about the time-course of speaker information when the speaker is a member of a potentially **stereotypable** group that listeners have a lot of experience with.

The Question

How quickly do listeners incorporate information about speaker sex when hearing words that exhibit a sex bias?

Visual world paradigm



"Click on the **beaker**"

(Allopenna et al, 1998)

The Prediction

Listeners will look more quickly to targets when spoken by a speaker that matches the target's sex bias.

Stimuli and Procedure

Sex bias in two corpora:
- **Switchboard** phone calls between strangers; predetermined topics (Godfrey & Holliman 1997)
- **SpeedDate** in-person conversations between strangers; open topic (Jurafsky et al 2009)

We define *female bias* of a word as freq(word | female_speaker) / freq (word | male_speaker).
10 cohort competitor pairs chosen to maximize sex bias and imageability. (+ 5 neutral pairs, 15 fillers)

male-biased words

word	avg female bias
bullet	0.65
whistle	0.26
collar	0.35
wallet	0.15
hardware	0.13
tea	0.48
hand	0.64
cave	0.00
stadium	0.50
vest	0.00

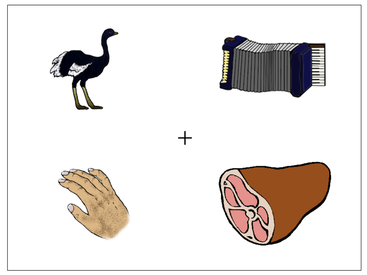
female-biased words

word	avg female bias
book	1.213
witch	inf
college	1.575
water	1.549
heart	2.201
teacher	1.714
ham	2.256
cake	4.837
steak	4.055
vegetable	3.609

eye-tracking trials

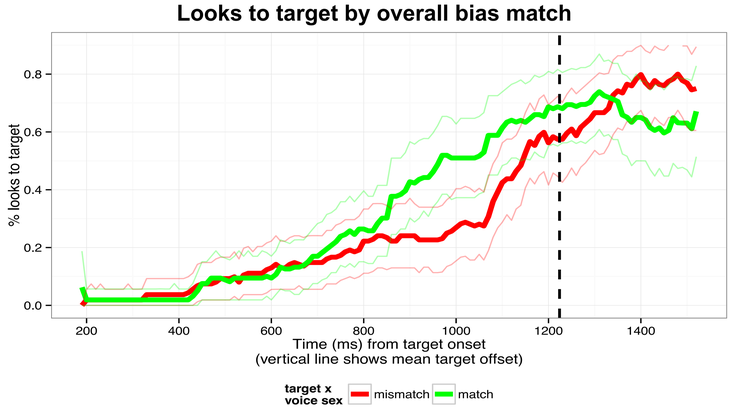
word	speaker	type
hand	M	match
hand	F	mismatch
ham	M	mismatch
ham	F	match

- trial types counter-balanced across four lists
- critical pair breakdown:
  - 1/3 m-bias
  - 1/3 f-bias
  - 1/3 neutral



Results

Looks to target by overall bias match



Looks to target increase between about 800ms and 1200 ms **when target bias matches speaker sex.**

The overall difference in this window is significant ($\chi^2(1) = 385.7$, $p<0.001$)

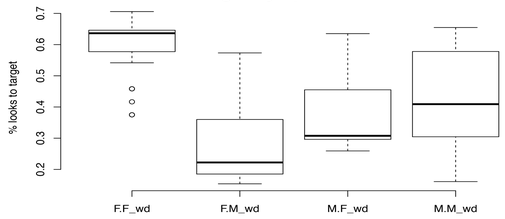
Listeners look faster to images when speaker sex matches word sex-bias.

The benefit for word-speaker match occurs **regardless of speaker or word sex.**

logit regression: looks to target by word bias and speaker in window:

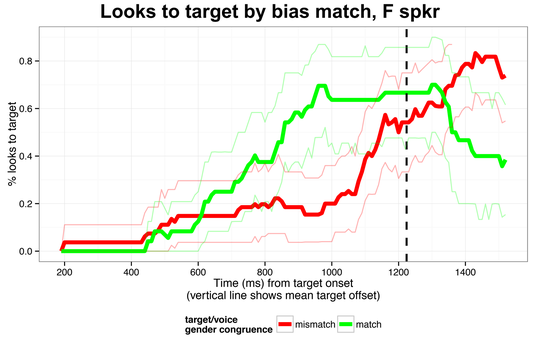
intercept (spkr=F, bias=F)	0.40	***
spkr=M	-0.97	***
bias=M	-1.38	***
spkr=M*bias=M	1.58	***

Looks to target by speaker and word sex

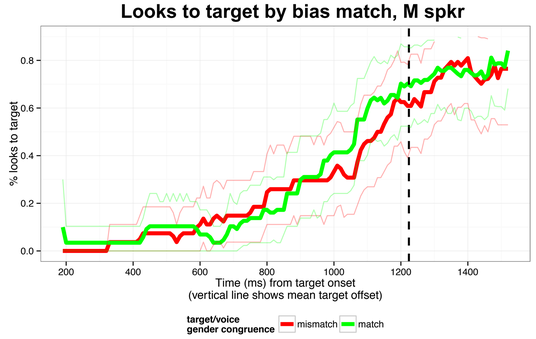


But the effect seems to be **driven by the female speaker**

Looks to target by bias match, F spkr



Looks to target by bias match, M spkr



Discussion

Some words in corpora show **evidence of sex-bias**: they are used more often by either men or women.

This sex bias in usage is **reflected in eye movements** during spoken word recognition.
- Listeners look faster to images when the sex of the person naming the image matches the sex bias of the name

---

Consistent with an **exemplar account**
- people hear "ham" spoken more often by women
- the string [hæ] spoken by a woman activates more exemplars of "ham" than of "hand"
- listeners look to the "ham" faster based on this increased activation

---

Provides a **fine-grained look** at the incorporation of sex information into spoken word processing
- overall effect does not emerge until after 800ms, so does not contradict the idea that this information is incorporated late
- but, for female speaker, looks start to diverge around 600ms

---

Sex bias match benefit appears stronger for female speaker than for male speaker.
- possibly due to overall faster looks to female-biased target
- may be because female-biased targets are shorter, more frequent, or more imageable

May suggest a model in which women's speech is **marked** relative to men's.
- male behavior is considered the default, while female behavior is implicitly or explicitly marked

(Eckert & McConnell-Ginet,(2003, Ch 1)

- supported by recent work, in which semantic associations are stronger and more predictive in online tasks when the speaker is a woman

(King & Sumner, 2013)

References

Eckert, P. and S. McConnell-Ginet (2003). *Language and Gender*. New York: Cambridge University Press.
Godfrey, J. J. and E. Holliman (1997). Switchboard-1 Release 2. Linguistic Data Consortium, Philadelphia.
Hay, J. & A. Walker (2013). Skewed experience with words affects lexical access patterns. Paper presented at Variation and Language Processing (VALP) 2. Christchurch, January 2013.
Johnson, K., E. Strand, and M. D'Imperio (1999). Auditory–visual integration of talker gender in vowel perception. Journal of Phonetics, 27(4):359-384.
King, E. and M. Sumner (2013). Voice-specific lexicons: acoustic variation and semantic association. Paper presented at New Ways of Analyzing Variation (NWAV) 42. Pittsburgh, PA, October 2013.
Luce, P. A., C. McLennan, and J. Charles-Luce (2003). Abstractness and specificity in spoken word recognition: indexical and allophonic variability in long-term repetition priming. In J. Bowers & C. Marsolek (Eds.), *Rethinking implicit memory* (pp. 197–214). England: Oxford University Press.
McLennan, C. and P. A. Luce (2005). Examining the time course of indexical specificity effects in spoken word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31 (2): 306-321.
Strand, E. and K. Johnson (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (ed.), Natural language processing and speech technology: Results of the 3rd KONVENS conference, Bielefeld, October 1996 (pp 14-26). Berlin: Mouton
Sherman, J. (1996). Development and mental representation of stereotypes. Journal of Personality and Social Psychology, 70(6): 1126-1141.

Questions? E-mail Ed (etking@stanford.edu)

This material is based upon work supported by the National Science Foundation Grant No. 1226963 to Meghan Sumner.
Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.