

# Role of prototypicality and frequency on phoneme processing

Bridget J. Smith (Ohio State University)



## Contrastiveness of eth /ð/ and theta /θ/

- In OE, /ð/ was allophone of /θ/ between voiced sounds, but became contrastive in ME
- Modern /θ/ and /ð/ maintain a similar distribution to historical allophonic distribution
- /θ/ and /ð/ share common characteristics, and have an overlapping distribution
- Voicing is predictable from environment (Smith 2007)
- Word-initial /ð/ only in function words; word-initial /θ/ in content words
- Consonant confusion matrices show them as barely confusable (Miller and Nicely, 1955), or moderately similar (Mohr and Wang, 1968)
- The contrast carries a low functional load, distinguishing only 13 possible minimal pairs



## Prototypicality

- /θ/ is the primary sound correspondence for <th>, based on anecdotal evidence
- Loanwords with orthographic <th> generally take /θ/, (ex: thermos, arthritis)
- Based on production from conversation and lab speech, prototypical /θ/ could be described as a primarily (inter)dental, primarily voiceless fricative
- It is unknown which form best represents /ð/ because of the wide amount of variation, reduction, and uneven distribution in production

## Frequency



- /θ/ has a higher type frequency than /ð/, both overall and in each position
  - /θ/ initial: 196 medial: 408 final: 111 (Brown corpus)
  - /ð/ initial: 33 medial: 128 final: 10
- /ð/ has a higher overall token frequency from a few high-frequency function words
  - /θ/ initial: 5,367
  - /ð/ initial: 85,278

## Research Question

1. Given the lack of contrast in production and lack of functional contrast, what is the perceptual status of the voicing contrast between /θ/ and /ð/?

## Experiments

- 2 experiments were conducted to try to determine if eth and theta are perceptually contrastive, and if they are similarly contrastive to /s/ and /z/
- Phoneme monitoring
    - Do listeners respond equally quickly to both phonemes? - gets at dominance
    - Can listeners reliably distinguish between two phonemes? - gets at perceptual contrastiveness
  - (Word) Identification and Discrimination
    - Can listeners reliably distinguish between two words in a minimal pair? gets at functional perceptual contrastiveness in words

## Experiment 1

- 24 participants: 7 men and 17 women, between the ages of 19-32
- Participants were instructed in each block to listen for one of the following: the sound of \*s\* as in *sun*, the sound of \*z\* as in *zen*, the sound of \*th\* as in *thin*, and the sound of \*th\* as in *then*.
- Response times indicate how quickly they responded when they heard the correct sound

Table 1. Example strings for the theta block in the phoneme monitoring task

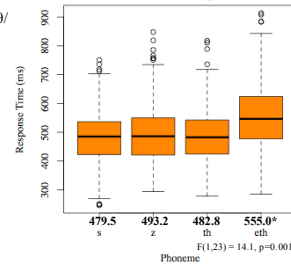
type of string	phoneme order – target and catch syllables in bold						
target	ha	ra	bæ	mæ	<b>θi</b>	ga	tæ
catch	<b>f</b> a	ni	<b>ð</b> i	da	sæ	ræ	zi
filler	da	fi	wa	ta	ga	di	læ

## Experiment 1

Table 2. Accuracy rates by target block

Target block	Hit rate %	Correct rejection %	Overall accuracy %	Correct rejection minus /θ/, /s/, and /ʒ/ %	Overall accuracy minus /θ/, /s/, and /ʒ/ %
th	94.1	43.1	66.6	90.5	92.3
eth	79.5	63.6	76.8	92.8	86.2
s	95.5	81.6	88.4	94.6	95.1
z	97.9	97.6	97.7	97.6	97.7

## Phoneme Monitoring Task RT



- Hit rate for /ð/ significantly lower than /θ/ (t(23)=4.1279, p<0.001)
- Error rate on catch trials for /θ/ and /ð/ not significantly different than accuracy overall
- Response Time (RT) longer for /ð/ (t(23) = -7.9, p<0.001)

## Experiment 2

- 25 participants: 8 men and 17 women, between the ages of 18-22
- 7-step voiceless to voiced continua using a synthesis of *sue* and *zoo* and *thigh* and *thy*.
- Two continua for each phoneme pair, based on transition into following vowel.
- Identification - Participants were played one sound at a time and asked to decide if the word was \*thigh\* or \*thy\* for the th- blocks or \*sue\* or \*zoo\* for the s-z blocks.
- Discrimination - stimuli were presented in 2-step AX pairs. Participants answered 'same' or 'different'.

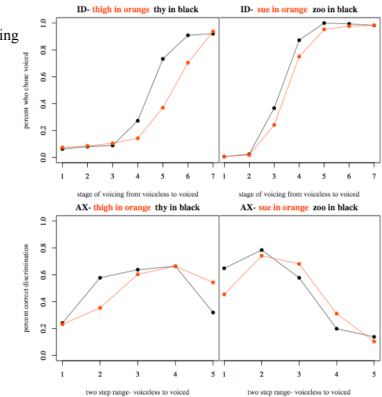
## Word frequency

- *thigh* and *thy* occurred 17 and 7 times out of 1.02 million in the Brown corpus respectively
- *sue* and *zoo* occurred 20 and 10 times respectively (Kucera and Francis, 1967)

## References

- Bybee, Joan. 1985. Morphology: A study into the relation between meaning and form. Amsterdam: John Benjamins.
- Bybee, Joan. 2001. *Phonology and Language use*. Cambridge University Press.
- Kucera, Henry, and W. Nelson Francis. 1967. *Computational Analysis of Present-Day American English*. (The Brown Corpus). Providence: Brown University Press.
- Miller, George and Patricia Nicely. 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27(2): 338-352.
- Miller, Joanne L. 2001. Mapping from acoustic signal to phonetic category: Internal category structure, context effects, and speeded categorization. *Language and Cognitive Processes*, 16 (5/6): 683-690.
- Mohr, Burkhard and William S. Y. Wang. 1968. Perceptual distance and the specification of phonological features. *Phonetica* 18: 31-45.
- Smith, Bridget J. 2007. The seeds of sound change don't fall far from the tree. Unpublished ms, presented at OSU Colloquium Fest, Spring 2007.

## Results



## Identification

- More tokens of greater voicing were judged to be instances of /θ/ than /s/ (F(1,40)=263.3, p<0.001)
- Transition information was significant for /θ/ and /ð/ (F(1,42)=15.4, p<0.001) but not for /s/ and /z/

## Discrimination

- /θ/ and /ð/ boundary is significantly to the right of the boundary for /s/ and /z/ (F(1,24)=187.3, p<0.001)
- No significant difference in total percent discriminated

## Discussion

1. To answer the research question, /θ/ and /ð/ are as fuzzy in perception as in production
  - High error rates in exp. 1 show poor phoneme discriminability.
  - Exp. 2 shows some discriminability for minimal pair, indicating that there are two mental representations for the words *thy* and *thigh*
  - But strong bias for *thigh*, despite similar word frequency
  - Suggests type frequency or prototypicality of phoneme plays a role



## Prototypicality

- Non-prototypical stimuli that are closer to a phonemic boundary may cause longer response times (Miller 2001)
- More difficult (longer RT) to choose /θ/ as a non-prototypical example of /ð/ than it was to choose /ð/ as a non-prototypical example of /θ/
- Identification is biased toward /θ/ as conceptually primary representation of <th>
- When one sound dominates the conceptual space, and the other sound may be classified by language users as an instance of the dominant sound, the relationship is typically one of allophony.



## Frequency

- Slower RT for /ð/ in exp. 1 suggests weaker mental representation
- Bias toward /θ/ suggests more dominant representation
- Repetition can lead to strength of representation (token frequency), or reduction (type frequency)- Bybee 1985, 2001)
- Preference for /θ/ could simply be due to higher type frequency
- Difficulty processing /ð/ may be a result of huge variation in huge number of tokens, making matching difficult

## Other

- Difficulty in discrimination and processing could be due to other artifacts of the acoustic signal