Eth and Theta: A tale of two phonemes

Bridget Smith

1 Introduction

In the later stages of Old English, eth /ð/ was an allophone of theta /θ/, found intervocalically, as is true of the other voiced-voiceless fricative pairs at the time (Lass 1994: 71-72). This voicing distinction became phonemic in Middle English during a period that contained intense borrowing, reduction of the inflectional paradigm, and unstressed syllable reduction and loss. (Lass 1992: 58-9). This shift is evident through spelling changes, as <z> and <v> were introduced as graphemes. Eth and theta were unique in that around the time of this spelling reform, both the thorn <þ> and the eth <ð> characters dropped out of usage completely, and were replaced by a digraph <th> to represent both the voiced and voiceless dental fricative (Hogg 1992: 76-7). It is therefore difficult to guess whether they participated in the phonologization at the same time as /f/ and /v/ or /s/ and /z/, or if it occurred later. Due to the lack of loanwords containing a voiced dental fricative during this time, it is reasonable to guess that speakers had less motivation to phonologize the voicing distinction between theta and eth, and that the pattern of intervocalic voicing for <th> may have continued for a long time.[1]

As many word-final syllables were reduced and eventually lost, especially in the verb inflections, eth became a word-final phoneme in some verbs that were derived from nouns (teeth~teethe). This created a contrastive environment in final position with minimal pairs consisting of noun-verb alternations. Although this contrast could have paved the way for more voicing contrasts in word-final <th>, there are none. The word-final voicing contrast in the dental fricative seems only to apply to these noun-verb pairs, and the voicing can be seen as serving a grammatical purpose, rather than a lexical one. In modern usage, however, most of the instances of verb-final –th, such as in mouth or sheathe, can be pronounced as voiceless, so it seems that not all speakers recognize the voicing contrast as a meaningful one.[2] Word-internally, there is little room for contrast, with inflected forms having eth (for example, *northern*) and compounds maintaining theta (*northwest*), while loan words, presumably borrowed monomorphemically, generally contain theta (*ether*). The main contributor to phonologization occurred around the 14th century when the word-initial <th> of a handful of high-frequency function words was re-analyzed as being primarily voiced due to its reduced nature (Lass 1992: 59), thus allowing a more salient contrast in word-initial position, though this remains a contrast only between function words and content words, which may also be considered morphological.

In modern English, the voiced and voiceless dental fricatives are considered to be two separate phonemes. Before we address theta and eth specifically, some terminological distinctions are in order. Traditionally, phonemicity is treated as a binary classification. Either a sound is a phoneme, or it is not. Sounds that are interchangeable and which do not contrast are

---

[1] That is to say, that regardless of one's conception of phonologization, if there is no environment in which the sounds might contrast, there is not necessarily reason to believe that speakers perceive a contrast, unless there is evidence to the contrary.

[2] That is to say, the salience of the noun~verb voicing distinction is not very strong among many modern speakers, regardless of whether this is due to analogical processes, reconnecting the verb with the noun, or the increased productivity in the modern language of "verbing", that is the creation of a verb from a bare noun stem with no morphological changes besides the regular verbal inflections, such as *google* (noun): *to google* (verb), *googling, googled*, etc.

allophones, and sounds which contrast in any environment must exist in a phonemic relationship. However, in recent times, we have come to realize that there are a variety of relationships between and among sounds that have varying degrees of contrastiveness (see Goldsmith 1995, Hualde 2005, Scobbie & Stuart-Smith 2006, 2008, among others; and especially Hall 2009 for an in-depth assessment of this problem). In this paper, many of the criteria for determining contrastiveness are examined for theta and eth, but while applying many of the traditional tests, I maintain a position that the notion of phoneme does not necessarily entail a binary distinction, but that it may, in fact, be gradient. If we were to assume that any degree of contrast must entail a phonemic relationship that is equal in contrastiveness to any other phonemic relationship, we would be missing some of the most interesting details about the relationship between theta and eth. It is hoped that even the most devout traditionalist will agree, by the end of this paper, that the relationship between eth and theta, at the very least, problematizes the traditional concept of phoneme.

The contrast between theta and eth in American English carries virtually no functional load, having only thirteen possible minimal pairs, as shown in table 1. Seven of these have alternate pronunciations which can render them non-contrastive, such as mouth~mouth (the verb with eth or theta) or ether~either ([iðɚ] or [aɪðɚ]). Of the remaining six, two are noun-verb pairs which do not contrast in any syntactic position, and are derivationally related (teeth~teethe, wreath~wreathe), one is composed of an archaic form which appears only in compounds (sooth~soothe), and the only two pairs with the potential for contrast in real usage (thigh~thy, thou.~thou), contain forms of an archaic pronoun that are only used in formulaic utterances.

Table 1: minimal pairs for eth and theta, all positions

Minimal pair	**Part of speech**
birth/birth (both pronunciations possible in verb)	noun / verb
ether/either (except when either is [aither])	noun / adj,pro,conj,
mouth/mouth (both pronunciations possible in verb)	noun / verb
froth/froth (both pronunciations possible in verb)	noun / verb
loath/loathe (both pronunciations possible in noun)	noun, adj / verb
oath/oathe (both pronunciations possible in verb, archaic)	noun / verb
sheath/sheathe (both pronunciations possible in verb)	noun / verb
sooth/soothe (archaic - used only in compound)	noun, adj / verb
teeth/teethe (verb usually only used in –ing)	noun / verb
thistle/this'll (multi-word contraction)	noun/ pro+verb
wreath/wreathe	noun / verb
thigh/thy	noun / pro
thou. (clipping of thousand)/thou	number/pro

The allophonic relationship in Old English between voiced and voiceless fricatives is a common result of intervocalic lenition, which is a well-documented phenomenon in many languages, most frequently affecting voicing and sonority. For example, in Spanish, intervocalic voiceless stops became voiced stops, and then voiced fricatives (Lat. vita > Sp. vida > [viða]). Stops may become fricatives, voiceless stops or fricatives may become voiced, while voiced fricatives may become approximants. There is a phonetic basis for this kind of reduction. As reported for /s/ and /z/ and /f/ and /v/ (Cole and Copper 1975), reduced duration in fricatives may create the perception of voicedness. This is refuted to some degree by Jongman (1989), who shows that even when reduced to a small slice of only the fricative (30ms to 70ms), most fricatives can still be fairly well distinguished in voicing, though not theta and eth, which require the full amount of duration and vowel information for proper identification (and even then had the lowest identification rates of all).[3] Stevens et al's (1992) acoustic analysis shows more specifically that listeners are attentive to transitions between fricatives and vowels, and are more likely to perceive a fricative as voiced if it contains even short periods of voicing, as long as the voicing is adjacent to the fricative-vowel boundary, thus maintaining the transition information. The duration of voicing that is necessary for perception of voicedness changes depending on the duration of the fricative, such that longer fricatives are more likely to be perceived as voiceless. Although they did not look at theta and eth, van Son and Pols' (1999) study on consonant reduction showed F2 slope difference in formant transitions were usually smaller in more reduced syllables. They also found slightly reduced duration, and a lower center of gravity and reduced amplitude difference for consonants in reduced syllables, which implies more strength in lower frequencies, as would result from less constriction and thus more strength in the lower harmonics and formant frequencies.

These results are all consistent with intervocalic voicing, but we generally do not expect these processes of reduction to take place word-initially, and in fact, the only word-initial voiced dental fricatives belong to a very small class of very high frequency function words, such as *the, then, this,* and *that*. Hay, Pierrehumbert, and Beckman (2004:38-39) write that function words such as *the*, "might be considered lexically specified as unstressed," which is more or less the common consensus of historical linguists of an earlier era (see, e.g., Prokosch 1939:83, who calls them "habitually unaccented words"). There is generally no pause or break between them and the surrounding words, and even when they occur at the beginning of a phrase, they have the least important information, and are usually one of the least stressed items. This low-stress, highly predictable position, and their high token frequency leave them able to be reduced in the manners described above (for a historical treatment, see Lass 1992. For a phonetic treatment of reduction, see Pierrehumbert 2000, Jurafsky, et al. 2000, Ernestus et al. 2006, among others). Though when they do occur in prosodically prominent stressed positions, they may be voiceless. They may also become voiceless when they assimilate to an adjacent voiceless segment (Smith 2007).

New words that are adopted into English containing orthographic <th> will generally be pronounced with theta, whether word-initial, word-internal, or word-final, as can be seen by glancing through a dictionary (for example: *thegosis, anorthic, shibboleth*). Possibly because the orthographic <th> frequently represents an aspirated /t/ in the donor language (*Neanderthal, aesthetic*), words with word-final <th> are not borrowed as often, though when word-final <th> is adopted or created, it is expressed with theta (*math, mammoth*). Word-final eth only occurs in uninflected verb forms, and has a low type frequency (10 (non-variable) words in the Brown Corpus, compared to theta with 111 types). Word-initially, eth has a type count of only 33, all function words, but a very high token count of 85,278, whereas theta has the higher word-initial type count with 196 words in the Brown corpus, and token count of 5,367. Medially, theta has a type frequency of 408, mostly in compounds and borrowings, while eth has a type frequency of only 128, mostly in inflected forms. It is probably this overall higher type frequency that leads speakers to perceive theta as the unmarked, more prototypical form of <th> in every position, just as higher type frequency predicts greater phonotactic acceptability in new words (Hay et al.,

---

[3] Phoneme confusion studies have found theta and eth to be variably confusable, depending on the study: from hardly confusable (Miller and Nicely 1955) to moderately similar (Mohr and Wang 1968).

2004, among others). Campbell and Besner (1981) found that talkers generally pronounced spelled <th> in nonsense words with theta, and were reluctant to use eth in word lists, even when grouped with other function words. It was only when the non-word was placed in a sentence such as *I can see four cans over thuz near a tree* that subjects could be brought to produce eth word-initially, and even then, there was a great deal of variation, with a large number of voiceless and indeterminately voiced tokens. It has been argued that word-initial eth bears the hallmarks of a morpheme which marks a certain closed class of specific function words, or at the very least, that speakers may derive from this class of words a sound-meaning correspondence between eth and the property of deixis (Janda 1985). Word-initial eth is not productive outside of this a closed class, and apparently not very strongly within it. Thus, morphologically, it may be considered a marked form. This coincides with the assessments of Mohr and Wang (1968), Chomsky and Halle (1968), among many others, who treat the voiced member of a voiced-voiceless pair as the phonologically marked form. The more recent discussions of frequency are comparatively similar, in that the unmarked form is more likely to have a higher type frequency, whereas the marked form will have fewer overall members, but these may have a higher token frequency (Albright and Hayes, 2003, Hay, et al., 2004, among others). Thus, eth is more likely to be the less productive and more marked of the two dental fricative sounds.

In production, the voiced and voiceless dental fricative phonemes occupy an overlapping region of sound across multiple dimensions of voicing, place, and manner. They are heavily influenced by the phonetic environment in which they occur. In conversational speech, the traditional criteria for distinguishing the voiced from voiceless fricative don't apply to eth and theta. Duration and amplitude, even relative amplitude, vary greatly, and voicing is determined more by prosodic position, stress, and voicing of neighboring segments (Smith 2007). Eth may be deleted, reduced to a vowel or an approximant, or hardened to a stop, de-voiced, fronted or, more commonly, retracted. Theta may be fronted or retracted, it may be voiced or turned into a stop, but is only rarely reduced to the level of an approximant or vowel. Some of this is certainly an artifact of eth and theta's location in words and phrases. As mentioned, eth occurs mostly (type-wise) intervocalically, and when it is word-initial, it is often not in a position of prominence, but rather in a reduced, high-frequency function word. Theta occurs primarily word-initially, in content words, which are more likely to occur in positions of prosodic prominence, in low-frequency borrowings, or at the boundaries of words and compounds.

The voicing contrast in the dental fricatives is treated as phonemic, but there is high predictability in spite of minimal pairs. The phonemic contrast in word-initial position can be predicted syntactically, the contrast in word-final position morphologically. The phonetic realization can be predicted to a high degree by prosodic position and stress, assimilatory voicing, and frequency-related reduction. The status of the contrastiveness between eth and theta has not been determined perceptually. Does their high predictability and phonetic overlap cause their mental phonemic representations to overlap in a similar way to their production, and, if so, to what extent? Are they perceived as members of one category or two? How do they compare to other, seemingly more contrastive, voiced-voiceless fricative pairs, such as /s/ and /z/? These are the research questions addressed by this paper.

Predictions about perception of the marked versus unmarked form vary. Mohr and Wang (1968) predict that the marked form should be easier to perceive because listeners will be more attentive to the presence of the marked feature rather than its absence. The discussion of frequency and neighborhood density has more or less usurped the conversation about markedness, but the marked form loosely correlates to lower type frequency; forms with a higher neighborhood density (type frequency) have been shown to have an inhibitory effect in lexical decision and naming tasks, presumably because they activate more competitors, so it takes longer to narrow down the playing field (Hutzler et al., 2004). Both hypotheses would predict slower response times for /s/ and theta, and faster response times for /z/ and eth. It is unknown what effect, if any, the high token frequency of eth should have on response times. Wagner and Ernestus (2007) showed that competitors across the entire phoneme class (fricatives, vowels, etc.) have an effect such that, the more segments in the inventory of a sound class, the slower response times become in a phoneme monitoring task. This may be confounded with a higher type frequency, though. The prediction, if carried to the next logical step, should be that sounds with closer competitors along continua of place, manner, and voicing should slow response times in a similar manner to the high neighborhood density effect, so that /s/ and /z/ may be processed more slowly than theta and eth because there are more alveolar sounds that may compete, while there are no other dental competitors; however, /s/ and /z/ are much more acoustically salient than eth and theta, so this effect may be cancelled out. Also, the dental fricatives and alveolar fricatives may stand in competition with each other.

Two more factors should be taken into consideration. One is that eth, being a greatly variable sound and thus having a larger perceptual space, may produce longer response times and greater error rates, similar to the effect found for vowels (e.g., Cutler et al. 1990, van Ooyen et al. 1991). In this case, theta may also have slightly longer RTs and greater error rates than /s/ and /z/, but not nearly as great a difference as eth would show. Another factor includes recent work that is trying to ascertain whether listeners use underlying representations (e.g., Pitt 2009) or explicit representations of surface productions, when accessing the phonemic shape of a word (e.g., Connine 2004). The "mediated access theory," seemingly favored in Pitt 2009[4] would suggest that if the surface form does not align with the underlying form, then an additional stage of mediation is necessary, and creates longer processing times, and generates less activation. The "direct access theory" favored in Connine 2004, would predict slower response times only in case the surface realization is not the most frequent realization. In case the mediated access theory holds, it makes the prediction that, if eth is a proper phoneme with a similar relationship to theta as /z/ has to /s/, and its underlying form is that of a voiced fricative, it should have proportionally related response times to theta as /z/ has to /s/. It would then also predict that if eth is an allophone of theta, that response times would be longer, because listeners would have to perform the extra step of "converting" to theta. The direct access account should predict longer response times for eth, regardless of its relationship to theta, because the most frequent surface realization of eth is greatly reduced, taking the form of an approximant, vowel-like segment, or being deleted.

A series of 3 experiments was conducted to determine if eth and theta are perceptually phonemic, if they are equally contrastive compared to the phonemic relationship between /s/ and

---

[4] Unfortunately, those involved in this debate (e.g., McLennan et al 2003, 2005; Connine 2004; LoCasto and Connine 2002; Janse et al. 2007, etc.) have shifted from studying segmental variants to deletions, which is problematic for more than one reason: orthography may play a very large role in creating the perception of the existence of a sound where none exists, i.e., phoneme restoration, especially if the word involved has a higher written frequency, or if the language user is highly literate. Unstudied also is the prestige weighting of orthographic effects, which may include spelling pronunciation. Another problem, addressed only as a footnote in Pitt 2009, is that in conversational speech, in some words with deleted variants, the sound may not actually be entirely deleted, and perceptual cues may exist for the "deleted" segment (e.g., Warner et al. 2009). Finally, other qualities found in reduced speech may confound the effects obtained by using non-reduced lab speech to represent words with deleted segments.

/z/, and to try to tease apart some of the details surrounding their phonemic, yet not-very-contrastive nature. The first experiment was a phoneme monitoring task, which shows a difference in accuracy rates between theta and eth and /s/ and /z/, especially regarding catch trials in which eth was mistaken for theta and vice versa, while the same did not hold for /s/ and /z/, and higher response times for eth. The second experiment included a discrimination task and an identification task, which show some kind of categorical perception is possible, but also reveal differences in the perception of eth and theta as compared to /s/ and /z/.

2 Methodology

2.1 Experiment 1-- Phoneme monitoring

2.1.1 Materials

A phoneme monitoring task was constructed in order to see how quickly and accurately listeners could respond to theta and eth and /s/ and /z/ for comparison. Stimuli were created using recordings produced by a trained male linguist. He was instructed to read the syllables slowly and carefully, using the same slight falling intonation on each word, and pausing between syllables. The recording was done in a sound booth, using a Shure model KSM27 cardioid condenser microphone with phantom power supplied through an Art brand Tube MP preamplifier, attached to a Macbook running the Audacity sound editing program. The linguist read a list of CV syllables from IPA transcription, including 19 phonemes that may begin a syllable in English, each paired with 5 vowels, /i, æ, u, a, aɪ/, for a total of 95 syllables, which were repeated three times each. The complete list can be found in the appendix. Only those syllables with /i, æ, a/ were used in the phoneme monitoring task, so that most syllables would not sound like a real English word, for a total of 57 syllables. 128 seven-syllable strings were randomly generated. Duplicate consonants in a string were replaced, as were any sounds within 2 features of the target phoneme that were located 2 syllables before the target in a string. The syllables were then concatenated into a string with 250 ms of silence in between each syllable. For each phoneme target, there were 15 target strings, composed of 6 filler syllables plus the target syllable, although those with the target in the first or last position were not included in the analysis (leaving 12-13 strings per phoneme for analysis). Each target phoneme's block also consisted of 14 seven-syllable filler strings, and 3 catch strings, which contained a syllable with the matched phoneme with the opposite voicing quality, paired once with each vowel, plus 6 filler syllables. An example of each type of string can be seen in table 2.

Table 2. 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	**θi**	ga	tæ
catch	ʃa	ni	**ði**	da	sæ	ræ	zi
filler	da	fi	wa	ta	ga	di	læ

Using the pairs of target syllables that were most similar to each other in intensity and duration to begin with, /s/ and /z/ and theta and eth were extracted and manipulated to match in amplitude and duration, so that any effect of phoneme would be not be due to longer or shorter duration or greater or lesser amplitude. The amplitude and duration of /s/ and /z/ were normalized to 71 dB and 125 ms, respectively. The amplitude of eth and theta was normalized to between 62-63 dB, and the duration was also held to a constant 125 ms. Small slices of the voiceless fricatives were removed to create matching duration, while the overall RMS intensity of each fricative was boosted or lowered across the board, so that the spectral slope particular to each segment would remain intact. The following vowels were also normalized to 76dB, though the adjustment was never more than two decibels. The length of the following vowels was not adjusted, but all target syllables fell into a range of 412-495 ms, those with /æ/ being generally the longest. The filler syllables were not manipulated; the overall RMS intensity of each syllable was within 62-81dB, and the overall duration of each syllable was within 347-594 ms.

2.1.2 Participants

Participants were drawn from a pool of introductory linguistics students who participated in the experiment in partial fulfillment of their course requirements. After excluding participants who reported speech or hearing disorders and those with a non-English background, data from 28 participants were included in the analysis. After analysis of the data, 4 additional participants with an overall error rate above 20% were also excluded. The remaining 24 participants were 7 men and 17 women, between the ages of 19-32 years old.

2.1.3. Procedure

There were 4 randomized blocks of 32 randomized trials, one block for each target, each block containing all 15 target strings, 3 catch strings, and 14 filler strings. Each subject heard each block once. Each block was prefaced by instructions and 3 practice trials. The instructions were presented on the screen, as follows: "You will be instructed to listen for a particular sound. You will then hear sequences of nonsense syllables. If you hear the sound that you have been instructed to listen for, press the SPACE bar."… "Do not press the SPACE bar if you do not hear the sound. The sound may not be in all of the sequences, and the sound you are listening for will change from time to time. Each time it does, you will be given new instructions."… "Listen for the sound of **s** as in the word sun (*z* as in the word zinc, *th* as in the word thin, *th* as in the word then.) Press the SPACE bar as quickly as you can, each time you hear the sound." Upon logging a response, the string was truncated, so that participants would know that their response had been recorded, and a new string began playing after a 500 ms pause. The strings were played over headphones at a comfortable listening volume, and participants' responses and response times were recorded in E-prime.

2.2 Results

Accuracy was recorded in terms of "hits," which indicate the proportion of targets correctly identified, out of all strings in which a target was located, and "correct rejections," that is, the proportion in which no phoneme was selected out of the filler trials. The accuracy rates for the dental fricatives are much lower in comparison to the alveolar sibilant fricatives. The accuracy rates are shown in table 2. The accuracy rate for eth was 76.8% and theta was 66.6%, and for /s/ was 88.4%, while /z/ was 97.7%. However, 83.9% of theta/eth false alarms stemmed from confusions with /f/ and /v/; theta was especially often confused with /f/. And 76.7% of the /s/ false alarms came from participants responding to syllables beginning with <sh> /ʃ/. After removing these trials, the overall accuracy rates for eth and theta combined (minus f/v), is 89.3%. The overall accuracy rates for s/z, excluding <sh>, improve to 96.6%. The accuracy rates for eth are lower than for the other phonemes, while theta has similar, though slightly lower rates as /s/ and /z/. The cause of eth's lower accuracy is primarily that the hit rate for eth is significantly lower than for theta (t (23)=4.13, p<0.001).

Table 3. Accuracy rates by target block

Target block	Hit rate %	Correct rejection %	Overall accuracy %	Correct rejection minus /f/, /v/, and /ʃ/ %	Overall accuracy minus /f/, /v/, and /ʃ/ %
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

Even more interesting is how participants responded to the catch trials. Table 4 shows the false alarm rate on catch trials compared to overall hit rate on the corresponding sound. The false alarms on the catch trials for /s/ and /z/ are fairly low, though /s/ was chosen in place of /z/ 16.7% of the time. Compared to the catch trials for eth and theta, this number is very small. The rate of false alarms for eth and theta during the catch trials is high enough to be an acceptable hit rate, but low enough to reveal that listeners may have been somewhat bothered by the phoneme mismatch. The hit rate and false alarms on catch trials is not significantly different for eth or theta. The difference is very significant for /s/ and /z/, as can be seen in the table below.

Table 4. Hit rates compared to false alarms on catch trials, arranged by which sound was played --target or catch. (Standard deviation in parentheses.)

Sound that was played -- target	Hit rate %	Sound that was played -- catch	False alarms- catch trials %	Paired t-test results
th	94.1 (0.12)	th (catch for eth trial)	86.1 (0.22)	t=1.66, df=23, p=0.1 *not significant*
eth	79.5 (0.24)	eth (catch for th trial)	76.4 (0.30)	t= 0.51, df=23, p=0.6 *not significant*
s	95.5 (0.08)	s (catch for z trial)	16.7 (0.22)	t=15.64, df=23, p<0.001
z	97.9 (0.04)	z (catch for s trial)	5.6 (0.16)	t=27.47, df=23, p<0.001

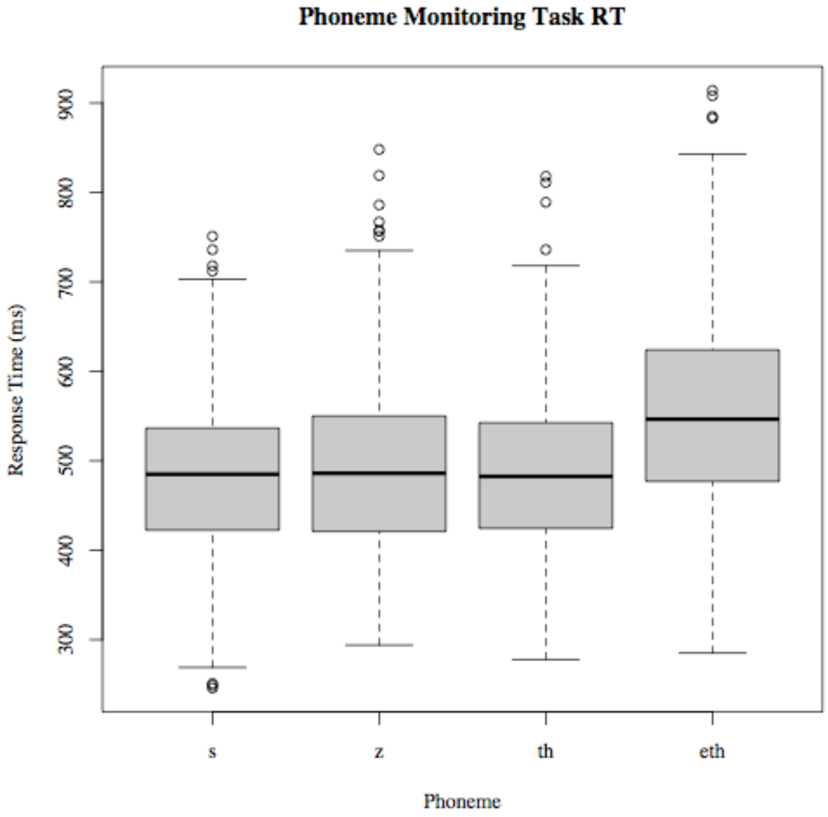


Figure 1. Response times for /s/ and /z/ and theta (th) and eth.

As with the accuracy rates, a difference in the relationship between eth and theta in comparison to /s/ and /z/ is evident; participants took much longer to respond to eth. Response times (RT) were measured from the onset of the fricative target. Each participant's response times were corrected for outliers by removing trials for which RTs fell below 100ms or were greater than three times the standard deviation from the mean of that phoneme for that participant. Those that fell above three standard deviations for the entire group were replaced by the average of the mean RT for that participant for that phoneme and the overall mean RT for that item across all participants. A repeated measures ANOVA (with place and voicing as within-subject factors) was performed on the response times for correct hits, which showed a significant main effect of place (F(1,23) = 25.3, p<0.001) and a significant main effect of voice (F(1,23) = 50.6, p<0.001), and a significant interaction between place and voice (F(1,23) = 14.1, p=0.001). Figure 1 shows median and interquartile ranges of RTs for each of the four target phonemes. We can see that the interaction in the ANOVA is caused by the much slower response times for eth than for the other phonemes. Planned comparison t-tests revealed a significant difference between response times for eth and theta (t(23) = 7.9, p<0.001) , and no significant difference between /s/ and /z/, or between /s/ and theta, or /z/ and theta.

2.2.1 Discussion

Eth and theta appear to be highly confusable, in spite of previous consonant confusion research, which shows them as barely confusable (Miller and Nicely, 1955), or only moderately similar (Mohr and Wang, 1968). In both of these studies, however, participants were extensively trained on each phoneme before completing the identification or similarity rating task, so we might take their results with a grain of salt. Training, of course, can lead to adjustments in perception (e.g., Kraljic and Samuel 2005, 2006), which may allow for improved discrimination, and can enable listeners to identify even phonemes that do not contrast in their native language (e.g., Logan et al. 1991, 1993, Lively et al. 1994).

In the current phoneme monitoring task, one might ask, were the instructions insufficient to eliminate orthographic effects? If this were the case, we might suppose that at least some of the listeners could have figured from the instructions, the different words used, and the juxtaposition of /s/ and /z/, that separate phonemes were meant for each block, but *all* subjects had *at least one* false alarm in the catch trials for eth ***and*** *at least one* false alarm in the catch trials for theta. If none of the participants could distinguish between eth and theta without extensive training, that might point to something more than just orthographic confusion. It might indicate that eth and theta belong not only to one grapheme, but also to one phoneme. For instance, we would not expect that listeners would require extensive training to tell the difference between the /k/ sound in *cat*, and the /s/ sound in *city*, though there might be occasional confusion, along the lines of what was found for the /ʃ/ sound represented by <sh>. This could be verified through further experiments.

The lower accuracy rates for eth (even removing the acoustically similar /f/ and /v/) than for theta or /s/ and /z/, point to greater difficulty processing this sound, and the lower RTs for eth also seem to indicate some kind of difficulty or complexity in processing, resulting in longer processing time. During the debriefing sessions, a number of participants independently commented that it seemed more difficult to listen for eth. There are multiple possible explanations for all of these results:

1) Although eth may be a phoneme in its own right, it may sometimes behave as a non-contrastive sound that is subordinate to theta (i.e., an allophone), which may take longer to process than the primary sound to which it serves as an allophone. Further experiments will need to be done to see if the results of phoneme monitoring are comparable to traditional allophonic pairs, such as /d/ and flap, or to other phonemes that double as allophones, such as schwa. Medial and final eth may also pattern differently, or may reinforce this notion.

2) In conversational speech, eth is rarely realized in an unreduced form as a full voiced fricative and is produced with a large amount of variation, depending on phonetic and prosodic context. Research on vowels shows greater difficulty in discrimination and identification when the vowels are taken out of context (e.g., Healy and Repp, 1982). This may be ascribed to the wide variability inherent to vowel production. Perhaps the slower response times to eth are an effect of the greater variability in the production of eth, which without being embedded in words or sentences, is difficult to determine out of context. Perhaps the great amount of variation in accumulated exemplars of eth makes it more difficult to decide membership in a group without lexical context. Although theta also displays great variability, it is much less than that of eth. The overlap in production may cause candidates with both phonemes to become activated, but the resolution is quicker for the smaller member of the group, that is the member with the least variation, theta.

3) Perhaps the voiced fricative is not a good prototype for the category. If most productions of eth are reduced to, say, approximants, the perceptual category may also have shifted in this direction. Non-prototypical stimuli that are closer to a phonemic boundary may produce longer response times (Miller 2001). This explanation is supported by the RTs for catch trials, as in table 5. The much longer response times when participants chose theta for eth suggest

that it was a more difficult decision to choose theta as a non-prototypical example of eth than it was to choose eth as a non-prototypical example of theta. In production, theta is more likely to be a fricative, whether voiceless or voiced, while eth is more often greatly reduced, though it may sometimes appear as a voiceless fricative. The difference in response times would also support a direct access theory interpretation (Connine 2004), even though words were not used in the experiment. Partial activation should still be stronger and faster for syllables beginning with the more frequently found variant. These RTs also suggest an unequal allophonic relationship, in which eth is a more common allophone of theta, but that theta is less often seen as an allophone of eth. This hypothesis could also be supported by evidence from production, and the historical relationship of eth and theta. This interpretation would be upheld by a mediated access approach, supposing an intermediary “th” phoneme. Further experiments are necessary to test each of these hypotheses.

Table 5. Average response time for hits (all phonemes) and catch trials (theta and eth). Standard deviation in parentheses. Hit rates were corrected for outliers, but there was not enough data, with only 3 catch trials per block, to perform corrections on catch trials (beyond an arbitrary 100ms lower cutoff point, and upper cutoff of 2900ms), which may account for some amount of deviation.

Sound that was played - target	mean RT in ms	Sound that was played - catch	mean RT in ms
th	482.8 (98.0)	th (catch for eth trial)	1086.1 (730.8)
eth	555.0 (120.7)	eth (catch for th trial)	757.8 (453.9)
s	479.5 (98.1)		
z	493.2 (104.8)		

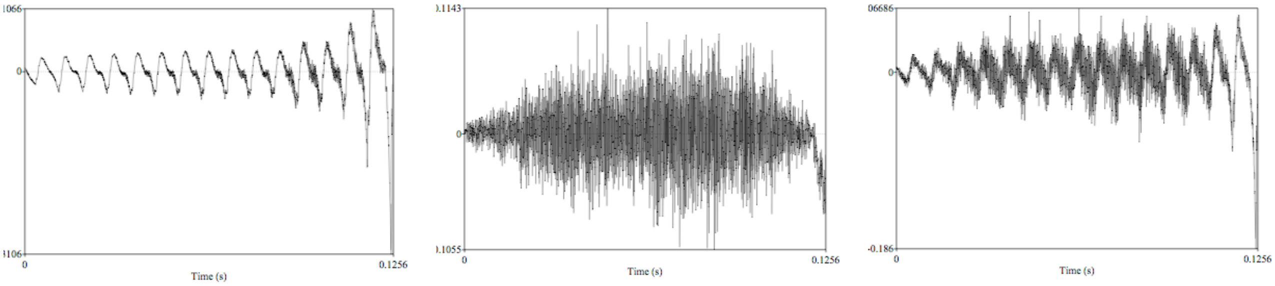
One thing is clear from the results of this experiment, that the relationship between eth and theta is not the same as the relationship between /s/ and /z/. The second experiment sought converging evidence for this observation by using the traditional test of categorical perception, an identification and a discrimination task, to examine the perceptual relationship between theta and eth from another perspective.

2.3 Experiment 2-- Identification and Discrimination

2.3.1 Materials

Continua were created using stimuli recorded from the first experiment, specifically the syllables /su/ ‘sue’ and /zu/ ‘zoo,’ and /θaɪ/ ‘thigh’ and /ðaɪ/ ‘thy’. These words were chosen because it is difficult to find a continuum for /s/ and /z/ at which neither end point creates a real word, while thigh~thy is the only word-initial minimal pair for eth and theta. It was thus decided that words should be compared against words in the spirit of the original study by Liberman et al. (1957). All the words are fairly low-frequency, and have a similar relationship between voicing partners, ‘thigh’ and ‘thy’ occurring 17 and 7 times, out of 1.02 million in the Brown corpus (Kucera and Francis, 1967) respectively, and ‘sue’ and ‘zoo’ occurring 20 and 10 times respectively.

The fricatives were removed from the beginning of each syllable, normalized for duration and amplitude, and then combined sample by sample, along a 7-step continuum. The end points for the th-continuum were the completely voiceless theta and the completely voiced eth. The end points for the s/z-continuum were the completely voiceless /s/ and the completely voiced /z/. The same algorithm was used for both continua, which combined percentages of each sample, in 16.667% increments. This process captures the differences in amplitude at each frequency and averages them within each sample of the waveform (44,100 per second). A diagram of the waveforms for eth and theta and spectral slices for the 50% step is shown in figure 2. The top row shows the overall effect on the waveform and the second row shows a spectral slice taken from the midpoint of each sound. In the spectral slice taken from eth, note the first peak in amplitude, which represents voicing. It is not present in the sample for theta, but is present to a lesser degree in the blended sample, which should give the auditory impression of "weak" voicing. Note also the peaks in the higher frequencies, representing frication, which are strongest in theta, very subdued in eth, but more strongly present in the blended fricative.



50% voiced eth + 50% voiceless theta = 50/50 blended th (step 4)

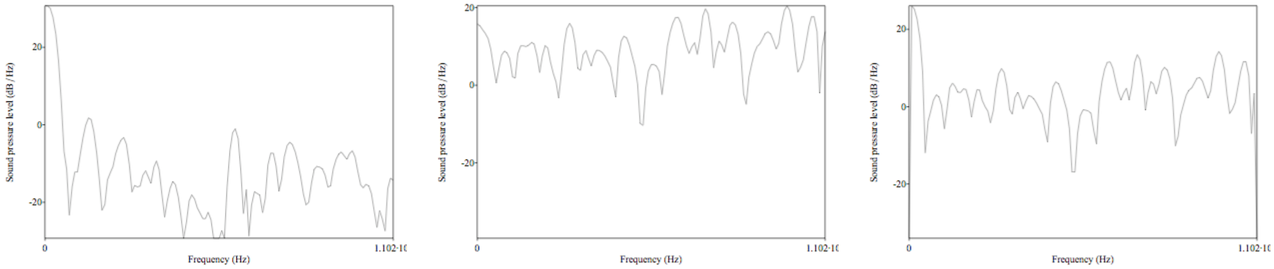


Figure 2. Blending of eth and theta at a rate of 50% each. In the top row the entire waveform of each fricative is shown along with the resulting blended fricative. In the second row, a spectral slice has been taken from the midpoint of each fricative to show how the blending process affects each sample in the waveform.

Five dental fricatives were created by this blending process (16.67% voiced, 33.33% voiced, 50% voiced, 67.67% voiced, and 83.33% voiced), and five alveolar sibilant fricatives. The endpoints were not manipulated, but retained the original acoustic characteristics. This created 7 points along a continuum of voicing and frication, such that the stronger the voicing, the weaker the frication and vice versa. Each fricative created in this manner was then reattached to each vowel from the words from which they were derived, creating two 7-step continua for each sound, one based on the vowel produced with the voiced fricative (zoo, thy) and one based on the vowel produced with the voiceless fricative (sue, thigh). Both vowels were used for each fricative continuum because it was thought that the transition information on these vowels might provide cues to voicing.

2.3.2 Participants

Participants were drawn from a pool of introductory linguistics students who participated in the experiment in partial fulfillment of their course requirements. All participants participated in both tasks. After excluding participants who reported speech or hearing disorders and those with a non-English background, as well as 2 participants who reversed the endpoints on the voicing continuum (that is, they seemed to associate eth with *thigh* and theta with *thy*), there were 25 remaining participants, 8 men and 17 women, between the ages of 18-22 years old.

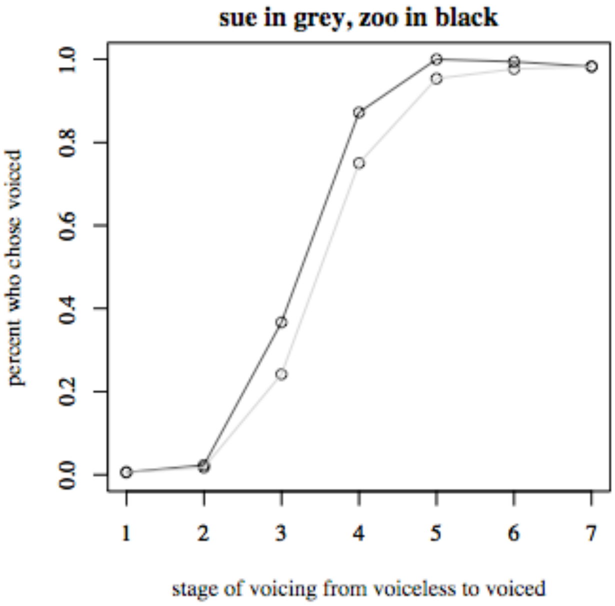
2.3.3 Procedure

The stimuli were used in an AX discrimination task and a following word identification task. The AX paradigm was used despite the usual preference for more difficult tasks such as 2AFC or ABX because it was thought the other formats might prove too difficult for eth and theta considering the high level of confusability displayed in the phoneme monitoring task.

First the stimuli for the AX discrimination task were presented over headphones, one at a time, in two randomized blocks, one for the s~z continua, and one for the theta~eth continua. Stimulus pairs were composed of matched pairs within the same continuum (0%~0%, etc.), 1-step pairs within the same continuum, in both orders of progression of voicing) (0%~16.7% and 16.7%~0%, etc.), and in two-step pairs in both orders of progression (0%~33.3% and 33.3%~0%, etc.). Presentation order of the stimulus pairs was also randomized. Each subject heard each stimulus pair 3 times. The instructions on the screen told participants to decide if the two syllables sounded the same or different. The sounds could not be replayed, but the experiment continued only after a selection was made.

The identification task was presented after the discrimination task in two randomized blocks in which the order of the stimuli were also randomized. Each subject heard each stimulus five times. The instructions on screen told the participants to select which word the syllable most sounded like, with the choices of sue and zoo for the s~z continua, and thigh and thy for the theta~eth continua. The sounds could not be replayed, but the experiment continued only after a selection was made.

2.4 Results



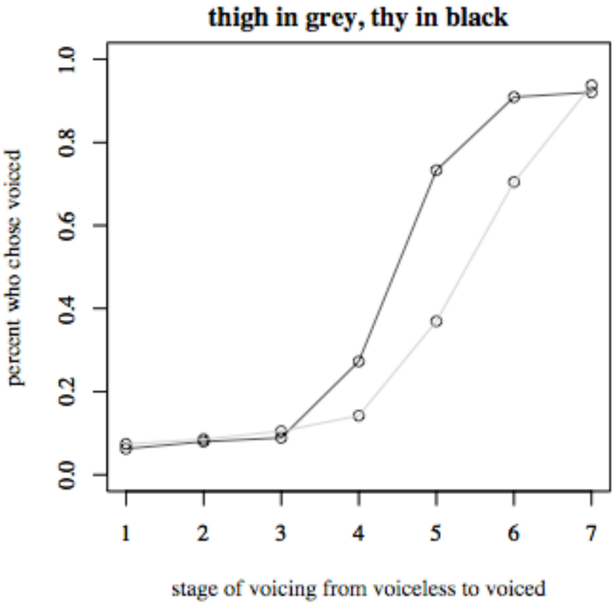


Figure 3. Results from the identification task: /s/ and /z/ on the left; eth and theta on the right. On the y-axis is percentage of "voiced" decisions; on the x-axis are the steps resulting from a 16.3% change in voicing from 0% voiced (voiceless) to 100% voiced (step 7). Black lines signify continua developed from originally voiced tokens (zoo, thy). Grey lines signify continua developed from originally voiceless tokens (sue, thigh).

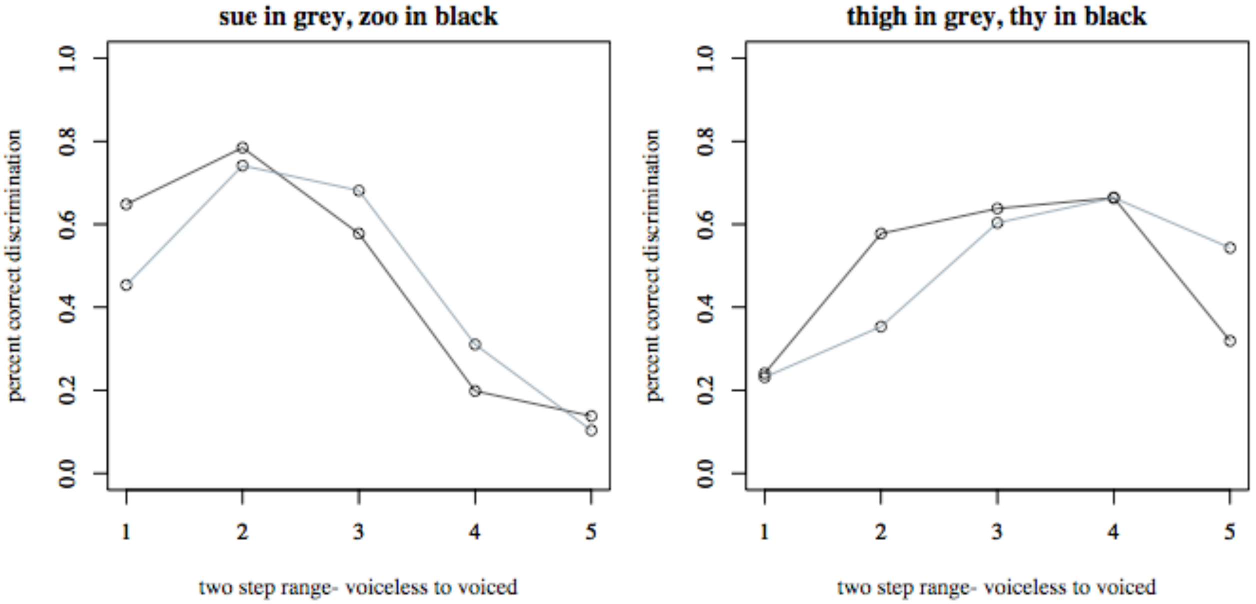


Figure 4. Results from the discrimination task: /s/ and /z/ on the left; eth and theta on the right. On the y-axis is percentage correctly identified as "different"; the x-axis shows the pairs of stimuli 2 steps apart from 0% ~ 33.3% voiced (step 1) to 66.7 ~ 100% voiced (step 5). Black lines signify continua developed from originally voiced tokens (zoo, thy). Grey lines signify continua developed from originally voiceless tokens (sue, thigh).

According to the theory underlying these tasks, sounds that vary along one continuum can be perceived categorically if they are members of two different phoneme categories. The boundaries from the identification task must align with the boundaries from the discrimination task. Eth and theta do show evidence of a perceptual categorical boundary, but the boundary is much closer to the voiced end of the continuum than the s/z boundary. The voicing judgments for the identification task were subjected to a 3-way ANOVA (with stimulus number, phoneme category (s/z or th), and continuum (voiced or voiceless) as within-subject factors), which showed a significant effect of stimulus number (F(1,40)= 1141.8, p<0.001), meaning voicing judgments differed based on which stimulus was heard. A significant main effect of continuum was found (F(1,39)= 63.2,p<0.001), showing that the responses differed based on whether the stimuli originated from the voiced or voiceless continuum. A significant effect of phoneme category was found (F(1,40)= 263.3, p<0.001), which we can see from the graphs, was due to a much lower number of eth/theta tokens being judged as voiced. There was also a significant interaction of stimulus number and phoneme category (F(1,41)= 14.3, p<0.001), phoneme category and continuum (F(1,41)= 9.82, p=0.003), and stimulus number and continuum F(1,40)=14.1, p<0.001), and a 3-way interaction of stimulus number, phoneme category, and continuum (F(1,42)=15.4, p<0.001). We can see the cause for the interactions with the continuum, as the voicing judgments of eth and theta are shifted even more toward the voiced side when the continuum using the originally voiced fricative is used. As previous studies have shown (Jongman 1989), transition information seems to be more important for the dental fricatives than for /s/ and /z/. Another ANOVA (with stimulus number and continuum as within-subject factors) was run on the eth/theta data to confirm that the difference between the two continua was significant. There was, in fact, a significant interaction between stimulus number and continuum (F(1,42)= 17.2, p<0.001). The same type of ANOVA for the s/z data alone did not show a significant interaction (F(1,42)=0.1, ns), meaning that which continuum the stimulus came from did not affect the responses for the s/z stimuli. It is noteworthy that even though the s~z boundary for the voiceless continuum is to the left of the boundary for the voiced continuum, it remains far to the right in comparison to both theta~eth boundaries, meaning that a dental fricative must be pretty strongly voiced, even with a voiceless transition, to be categorized as eth.

A 3-way ANOVA was also performed on the discrimination data, with similar results. The responses to the two-step intervals were analyzed because the discrimination of the one-step intervals was poor, barely reaching 40% at its peak. There was a significant main effect of stimulus pair (F(1,24)= 13.5, p=0.001), meaning difference judgments were different depending on which pair was heard. There was not a significant effect of continuum, meaning all 4 continua had approximately the same total percentage of correct discriminations. Phoneme category was also not significant, meaning that both s/z and theta/eth had approximately the same number of correct discrimination responses. There was a significant interaction between stimulus pair and continuum (F(1,24)= 10.8, p=0.003), meaning that there were differences in the percentage of correct discriminations for some of the stimulus pairs, depending on whether they were part of the voiced or voiceless continuum. As we can see, the judgments for the voiced continua do not completely overlay the voiceless continua judgments. There was also a significant interaction between phoneme category and stimulus pair (F(1,24)= 187.3, p<0.001), meaning that the percentage of correct discrimination responses differed greatly for the sue/zoo and thigh/thy stimuli based on stimulus number. This is due to the fact that the boundary for eth/theta is significantly to the right of the boundary for /s/ and /z/. There was no interaction of continuum and phoneme, meaning that, in this case, the overall percentage of correct discrimination responses based on which continuum was used (voiced or voiceless) did not differ significantly across the sue/zoo and thigh/thy sets. These results show that participants had about the same rate of success in discriminating between stimuli, whether sue and zoo, or thigh and thy; however the boundaries reflect the tendency shown in the identification task, for the boundary between eth and theta to be closer to the voiced end of the continuum than the s/z boundary.

2.4.1 Discussion

Both the identification and discrimination tasks showed evidence for perceptual categorical boundaries. The boundary for theta~eth was heavily biased toward the voiced side of the continuum, meaning that even in the word *thy*, a token had to be strongly voiced for listeners to identify it as eth. These results are inversely related to the observation made in the discussion section from the first experiment, where it was supposed that the large amount of variation that is normally observed in the production of eth might make it more difficult for listeners to identify eth. It made sense to suppose that hearing eth in the context of a word might make it easier to identify. This does seem to be the case because there is evidence for a perceptual boundary in the word identification and discrimination tasks; however, eth does not take up the greater perceptual space along the voicing continuum laid out in this experiment, but is chosen much less frequently than theta.

Because the stimuli used in the identification and discrimination task used similarly low-frequency words, it may be that activation of competing words containing both theta and eth are moderated by the higher overall type frequency of words containing theta, such that in times of doubt, the fallback is to go with the higher type frequency. Further experiments using non-words may shed light onto this hypothesis. If the hypothesis holds, it would also support the idea that if eth is more often processed as an allophone of theta, it will be more difficult to perceive out of context, as seems to be the case in the phoneme monitoring task. For example, Lipski (2006) found that native German speakers performed better at identifying the allophones of <ch>, that is /x/ and /ç/, when in a context they commonly occurred in, than when out of context. The unequal distribution of eth and theta may not allow them to be directly comparable, but it may be the case that theta may be identified more easily out of context than eth.

Identification and discrimination tasks are problematic in that they have been found to show categorical perception of non-linguistic stimuli (Cutting 1978), and allophones (Peperkamp et al. 2003), so they may not be as definitive at proving categorical membership as it was once thought. McMurray et al. 2003 offers a critique of the paradigm, suggesting that the traditional identification and discrimination tasks do not actually reveal an underlying phonemic identity. Rather, each sound has multiple acoustic cues, which may be weighted differently depending on acoustic and other cues, which, when manipulated, may cause "shifts" in perception. The identification and discrimination tasks, by design, focus only on one cue, or possibly a pair of cues that are inversely related, which may prove deceptive if other cues are not taken into account. Even if we accept the reigning hypothesis that categorical membership may be "proven" by this paradigm, it may also be possible for the results to show perception of categorical membership by the convergence of a boundary in both tasks while *only one* of the end points may be categorical. While tokens closer to category A may be perceived as members of category A, tokens on the other side of the boundary may be perceived as NOT A, rather than as members of a category B. To my knowledge, categorical stimuli that border on non-categorical stimuli have not been well researched. If this is one possible interpretation of the categorical discrimination and identification tests, it may be that eth is simply being categorized as NOT theta. Or it may well be that the voiceless fricative theta allows a good deal of voicing, and that the voiced fricative eth is just the beginning of the category for eth, and that if the continuum extended into a more sonorous, approximant-like version of eth, the border would lie more in the middle, where the category for theta overlaps into the voiced fricative region where eth begins. Further experimentation is necessary, and experiments using non-words may also show a different pattern.

3 Summary and Conclusions

The phoneme monitoring experiment revealed slower response times for eth than for theta. Catch trials, introducing eth where theta was the target, and vice versa, produced a very high number of false alarms, not significantly different from the hit rate. Accuracy rates for eth were lower than for any other phoneme. These results and the debriefing comments from participants suggest that it is a more difficult task to identify eth than theta, and that the two are highly confusable, which suggests that eth may at least sometimes be processed as an allophone of theta rather than a phoneme in its own right. The results also provide support for the hypothesis that while processing these syllables, lexical candidates containing both eth and theta are activated, as a result of the phonetic overlap between the two sounds in natural speech, though orthographic effects may also play a role. Longer response times for eth may be caused by its less contrastive nature, or because of its more variable range in production, and its tendency to be further reduced to an approximant.

Identification and discrimination experiments revealed a strong bias for theta, even when transition information from eth was left between the fricative and the vowel, and even though thy and thigh have roughly the same frequency as sue and zoo. This supports the idea that the target range for eth may be in the area of an approximant, but also leaves room for speculation about whether eth is categorically perceived, or if it perceived as an allophone of theta.

While this study creates an abundance of questions, there are a few points that it does support. The results of these experiments show that eth and theta are less contrastive than /s/ and /z/ in perception, which corresponds to their less distinctive nature in production. Participants do not process the relationship between eth and theta in the same way that they process the relationship between /s/ and /z/. Eth does seem to be the more marked form, though this seems to lead to greater difficulty in processing, and neither its higher token frequency nor the dimension of voicing gives an advantage in perception. The predictions that the unmarked phonemes with higher type frequency should be processed more slowly did not hold up in this experiment. It may be that near-ceiling performance masked potential differences between /s/ and /z/, but the differences between eth and theta go against predictions based on markedness and frequency. That there are fewer lexical competitors, within the phoneme category, as well as in neighboring dimensions of place and manner, should allow faster response times for theta and eth, but this is not the case, probably because of the greater acoustic salience of /s/ and /z/. Although /s/ has a higher type frequency and is the less marked of the s~z pair, it is not perceived as the default as theta is of the th-pair in identification and discrimination. Several directions for future research have been pointed out in the discussion of each experiment, which may shed more light on the mysterious relationship between eth and theta, and how less contrastive sounds are processed.

References

Campbell, Ruth and Derek Besner. 1981. This and THAP -- constraints on the pronunciation of new, written words. *The Quarterly Journal of Experimental Psychology, section A*, 33(4): 375-396.

Cole, Ronald A. and William E. Cooper. 1975. Perception of voicing in English affricates and fricatives. *Journal of the Acoustical Society of America*, 58(6): 1280-1287.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English.* New York: Harper & Row.

Cutler, Anne, Dennis Norris, and Brit van Ooyen. 1990. Vowels as phoneme detection targets. *Proceedings of the Internaional Congress on Spoken Language Processing.* Kobe. Japan. Vol. I, 581-584.

Cutting, James. 1978. There may be nothing peculiar to perceiving in a speech mode. In J. Reguin (ed.) *Attention and Performance VII.* Hillsdale, NJ: Erlbaum Associates. 229-243

Ernestus, Mirjam, Mybeth Lahey, Femke Verhees, and R. Harald Baayen. 2006. Lexical frequency and voice assimilation. *The Journal of the Acoustical Society of America*, *120*(2), 1040-1051.

Goldsmith, John. 1995. Phonological theory. In John A. Goldsmith (Ed.), *The handbook of phonological theory,* 1–23. Cambridge, MA: Blackwell.

Hall, Kathleen Currie. 2009. *A Probabilistic Model of Phonological Relationships from Contrast to Allophony*. Ph.D. Dissertation, the Ohio State University.

Hay, Jennifer, Janet Pierrehumbert, and Mary Beckman. 2004. Speech Perception, well-formedness, and the statistics of the lexicon. In John Local, Richard Ogden, Rosalind Temple (eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press. 58-74.

Healy, Alice and Bruno Repp. 1982. Context independence and phonetic mediation in categorical perception. *Journal of Experimental Psychology: Human Perception and Performance.* 8(1):68-80.

Hogg, Richard 1992. Chapter 3: Phonology and Morphology. In Richard Hogg (ed.), *The Cambridge History of the English Language. Vol I: The Beginnings to 1066*. Cambridge: Cambridge University Press. 67-167.

Hualde, Jose Ignacio. 2005. Quasi-phonemic contrasts in Spanish. In Vineeta Chand, Ann Kelleher, Angelo J. Rodriguez, and Benjamin Schmeiser (Eds.), *Proceedings of the 23rd West Coast Conference on Formal Linguistics*, 374-398. Somerville, MA: Cascadilla.

Hutzler, Florian, Jürgen Bergmann, Markus Conrad, Martin Kronbichler, Prisca Stenneken, and Arthur M. Jacobs. 2004. Inhibitory effects of first syllable-frequency in lexical decision: an event-related potential study. *Neuroscience Letters* 372: 179-184.

Janda, Richard D. 1985. An Autosegmental-Morphological Solution to the Problem of Phonesthemes and Blends. *11th Annual Meeting of the Berkeley Linguistics Society*, Berkeley, California: February 16-18, 1985.

Janse, Esther, Nooteboom, Sieb G. and Quené, Hugo. 2007. Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes* 22 (2): 161-200.

Jongman, Allard. 1989. Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America*, 85(4): 1718-1725.

Jurafsky, Daniel, Alan Bell, Michelle Gregory, and William D. Raymond. 2000. Probabilistic relations between words: Evidence from reduction in lexical production. In Joan Bybee and Paul Hopper (eds.), *Frequency and the emergence of linguistic structure.* Amsterdam: John Benjamins. 229-254.

Kucera, Henry, and W. Nelson Francis 1967. *Computational Analysis of Present-Day American English.* (The Brown Corpus). Providence: Brown University Press.

Lass, Roger. 1992. Chapter 2: Phonology and Morphology. In Norman Blake (ed.), *The Cambridge History of the English Language. Vol. II: 1066-1476.* Cambridge University Press. 23-155.

Lass, Roger. 1994. *Old English: A historical linguistic companion*. Cambridge: Cambridge University Press.

Liberman, Alvin M., Katherine S. Harris, Howard S. Hoffman, Belver C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology 54 (5).

Lipski, Sylvia C. 2006. *Neural correlates of fricative contrasts across language boundaries.* Ph.D. dissertation, University of Stüttgart.

Lively, Scott E., David Pisoni, Reiko A. Yamada, Yoh'ichi Tohkura, and Tsuneo Yamada. 1994. Training Japanese listeners to identify English /r/ and /l/ III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America.* 96 (4): 2076-2087.

Logan, John S., Scott E. Lively, and David B. Pisoni. 1991. Training Japanese listeners to identify English /r/ and/1/: A first report *Journal of the Acoustical Society of America.* 89: 874-886.

Logan, John S., Scott E. Lively, and David B. Pisoni. 1993. Training listeners to perceive novel phonetic categories: How do we know what is learned? *Journal of the Acoustical Society of America* 94: 1148-1151.

McMurray, Bob, Michael K. Tanenhaus, Richard N. Aslin, and Michael J. Spivey. 2003. Probabilistic Constraint Satisfaction at the Lexical/Phonetic Interface: Evidence for Gradient Effects of Within-Category VOT on Lexical Access. *Journal of Psycholinguistic Research.* 32(1): 77-97

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, Burckhard and William S. Y.Wang, 1968. Perceptual distance and the specification of phonological features. *Phonetica* 18: 31-45.

van Ooyen, Brit, Anne Cutler and Dennis Norris. 1991. Detection times for vowels versus consonants. *Proceedings of EUROSPEECH '91.* Genova. Italy. Vol. 3, 1451-1454.

Peperkamp, S., Pettinato, M., & Dupoux, E. 2003. Allophonic variation and the acquisition of phoneme categories. In B. Beachley, A. Brown, & F. Conlin (Eds.). *Proceedings of the 27th annual Boston University conference on language development* (Vol. 2). Sommerville, MA: Cascadilla Press. 650–661

Pierrehumbert, Janet. 2000. Exemplar dynamics: word lenition, frequency, and contrast. In Joan Bybee and Paul Hopper (eds.), *Frequency and the emergence of linguistic structure.* Amsterdam: John Benjamins. 137-158.

Prokosch, Edward. 1939. *A Comparative Germanic Grammar.* Philadelphia: LSA.

Scobbie, James M., and Jane Stuart-Smith. 2006. Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. *Queen Margaret University College Speech Science Research Centre Working Paper, WP-8.*

Scobbie, James M., and Jane Stuart-Smith. 2008. Quasi-phonemic contrast and the indeterminacy of the segmental inventory: Examples from Scottish English. In Peter Avery, B. Elan Dresher, and Keren Rice (Eds.), *Contrast in phonology: Perception and acquisition.* Berlin: Mouton.

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.

van Son, R. J. J. H. and Louis C. W. Pols. 1999. An acoustic description of consonant reduction. *Speech Communication* 28: 125-140.

Stevens, Kenneth N., Sheila E. Blumstein, Laura Glicksman, Martha Burton, and Kathleen Kurowski. 1992. Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters, *Journal of the Acoustical Society of America*, 91 (5): 2979-3000.

Warner, Natasha, Amy Fountain, and Benjamin V. Tucker. 2009. Cues to the perception of reduced flaps. *Journal of the Acoustical Society of America* 125 (5): 3317-3327.

Appendix: stimulus materials

all syllables recorded				
syllables used for experiment 1			syllables used in experiment 2 in bold	
ma	mi	mæ	mu	maɪ
wa	wi	wæ	wu	waɪ
pa	pi	pæ	pu	paɪ
ba	bi	bæ	bu	baɪ
fa	fi	fæ	fu	faɪ
va	vi	væ	vu	vaɪ
θa	θi	θæ	θu	**θaɪ**
ða	ði	ðæ	ðu	**ðaɪ**
sa	si	sæ	**su**	saɪ
za	zi	zæ	**zu**	zaɪ
ta	ti	tæ	tu	taɪ
da	di	dæ	du	daɪ
na	ni	næ	nu	naɪ
ra	ri	ræ	ru	raɪ
la	li	læ	lu	laɪ
ʃa	ʃi	ʃæ	ʃu	ʃaɪ
ka	ki	kæ	ku	kaɪ
ga	gi	gæ	gu	gaɪ
ha	hi	hæ	hu	haɪ