

# Sink positive: Linguistic experience with *th* substitutions influences nonnative word recognition

Adriana Hanulíková · Andrea Weber

Published online: 30 December 2011  
© Psychonomic Society, Inc. 2011

**Abstract** We used eyetracking, perceptual discrimination, and production tasks to examine the influences of perceptual similarity and linguistic experience on word recognition in nonnative (L2) speech. Eye movements to printed words were tracked while German and Dutch learners of English heard words containing one of three pronunciation variants (/t/, /s/, or /f/) of the interdental fricative /θ/. Irrespective of whether the speaker was Dutch or German, looking preferences for target words with /θ/ matched the preferences for producing /s/ variants in German speakers and /t/ variants in Dutch speakers (as determined via the production task), while a control group of English participants showed no such preferences. The perceptually most similar and most confusable /f/ variant (as determined via the discrimination task) was never preferred as a match for /θ/. These results suggest that linguistic experience with L2 pronunciations facilitates recognition of variants in an L2, with effects of frequency outweighing effects of perceptual similarity.

**Keywords** *th* substitutions · Spoken-word recognition · Speech comprehension · Nonnative word recognition · Eyetracking

---

A. Hanulíková · A. Weber  
Max Planck Institute for Psycholinguistics,  
Nijmegen, The Netherlands

A. Weber  
Donders Institute for Brain, Cognition, and Behavior,  
Nijmegen, The Netherlands

A. Hanulíková (✉)  
Basque Center on Cognition, Brain and Language,  
Paseo Mikeletegi 69,  
20009 Donostia, Spain  
e-mail: adriana.hanulikova@mpi.nl

When travelling on board German trains, German train conductors frequently make the English announcement *Ve vant to senk all pessengers for trevelling viz Deutsche Bahn*. This sentence contains several segmental deviations from standard English pronunciation, and travelers have to understand that, for example, the intended word behind *senk* is *thank*, and behind *viz* is *with*. But how do listeners recognize these intended words if they deviate so noticeably from the standard pronunciation? Does prior experience with the foreign accent influence how easily listeners can recognize the intended words, or is the perceived similarity between segmental deviations and the standard pronunciation more decisive? In the present study, we set out to empirically address these issues in two eyetracking studies in which listeners with different language backgrounds were presented with foreign-accented English words containing deviant pronunciations of the voiceless interdental fricative /θ/.

## Processing pronunciation variants

Spoken-word recognition is achieved by mapping the speech input onto lexical representations stored in the mental lexicon (see, e.g., Norris, 1994). One of the many challenges for this phonetic-to-lexical mapping process is the fact that spoken language is notoriously variable with regard to pronunciation. While some variations are contextually driven changes that result from adjusting sounds to the phonetic context, other types are idiosyncratic in nature due to the age, gender, and articulatory habits of speakers (see, e.g., Klatt, 1986; Kohler, 1991). As a consequence, the realization of a word in connected speech rarely resembles its canonical form—that is, the form provided by pronunciation dictionaries.

Previous research has suggested that the ease of word recognition depends on the nature of the variability

encountered in speech. This research has shown that variation resulting from regular phonological processes does not hinder word recognition. For example, assimilated word forms such as *hop* for English *hot* are recognized easily in assimilation-licensing contexts such as *hop bath* (Coenen, Zwitserlood, & Bölte, 2001; Gaskell & Marslen-Wilson, 1996, 1998; Gaskell & Snoeren, 2008; Gow, 2002). Similarly, schwa deletion (e.g., *histry* for *history*; Connine, Ranbom, & Patterson, 2008), flapping of *t* in English (e.g., *pretty* pronounced as /pɹɪfɪ/; Connine, 2004; Ranbom & Connine, 2007), resyllabification (e.g., the French word *orage* “storm” in the phrase *pe.ti.to.rage* “little storm”, where the dots indicate syllabification; Spinelli, McQueen, & Cutler, 2003), as well as context-dependent reduction of phonemes (e.g., the Dutch word *kast* “cupboard” pronounced as *kas*; Mitterer & Ernestus, 2006; Pitt, 2009) do not preclude word recognition.

In contrast to variation occurring in phonologically viable contexts, arbitrarily occurring phoneme substitutions in L1—that is, segmental deviations that do not result from L1 phonological processes—can disrupt word recognition (e.g., Marslen-Wilson & Zwitserlood, 1989), wherein word-initial substitutions constrain lexical access more than do word-final substitutions (Alloppenna, Magnuson, & Tanenhaus, 1998). In their cross-modal priming study, Marslen-Wilson and Zwitserlood showed that the Dutch word *honing* “honey” facilitated responses to the semantically related target *bij* “bee”, while the arbitrary substitution variant *foning* did not yield such an effect. The impact of segmental deviation on word recognition hinges on phonetic similarity—that is, the featural similarity between the deviant sound and the intended sound—as well as on perceptual similarity—that is, the perceived similarity between two phonemes. In general, similar substitutions disrupt word recognition less severely than do dissimilar ones, and dissimilar substitutions can even lead to a failure in word recognition (e.g., *corpora* is more similar to *gorpora* than to *horpora*; Connine, Blasko, & Titone, 1993; Connine, Titone, Deelman, & Blasko, 1997).

Recently, a number of studies have suggested that linguistic experience modulates the recognition ease of contextually driven variants in such a way that more frequent variants are recognized faster than less frequent variants (e.g., Connine, 2004; Mitterer & Ernestus, 2006; Pitt, 2009; Pitt, Dilley, & Tat, 2011; Ranbom & Connine, 2007; Ranbom, Connine, & Yudman, 2009). Ranbom and Connine, for example, have found significant priming effects as well as faster and more accurate lexical decisions for English words that are typically produced with a nasal flap than for words that are less often produced with a nasal flap. It has been further shown that the amount of experience with dialectal variants of one’s native language also modulates recognition ease. For example, British English listeners who have moved to the United States have learned to correctly interpret the medial flap in *todal* as /t/

(thereby recognizing the intended word *total*), but British listeners with little experience with general American pronunciations had more difficulties in recognizing the intended word (Scott & Cutler, 1984). Similarly, Sebastián-Gallés, Echeverría, and Bosch (2005) showed that native Catalan listeners who are frequently exposed to the Spanish dialect of Catalan have little difficulty recognizing Spanish-accented variants (e.g., \*/gəʎeðə/) as the intended Catalan words (e.g., /gəʎeðə/ “bucket”). However, atypical variants (e.g., /finestrə/ “window” pronounced as \*/finestrə/) are treated as nonwords (for effects of experience on dialect perception, see also Sebastián-Gallés, Vera-Constán, Larsson, Costa, & Deco, 2009; Sumner & Samuel, 2009).

So far, most studies of experiential effects in variant recognition have concentrated on variants that are contextually driven (e.g., Coenen et al., 2001; Gow, 2002; Mitterer & Ernestus, 2006; Pitt, 2009) or are based on word properties alone (e.g., Ranbom & Connine, 2007; Scott & Cutler, 1984) and that occur in one’s native language. Little is known about recognition of variants that are not driven by such local context and that occur in a nonnative language use. While native listeners have lifelong, abundant experience with variation in their L1, adult language learners have considerably less linguistic experience with their second language. Furthermore, their perception of L2 contrasts is often less precise than that of L1 listeners. It is therefore not clear how nonnative variants are recognized by L2 listeners. It is possible that L2 listeners do not primarily benefit from linguistic experience, but instead from perceptual similarity.

### ***Th* substitutions in foreign-accented speech**

An excellent test bed for investigations of the roles of experience and perceptual similarity is the occurrence of *th* substitutions in foreign-accented speech. The English voiceless interdental fricative /θ/, for example, presents a great deal of difficulty for many L2 learners. Several studies have not only described /θ/ substitutions across various L2 learners of English, but have also tried to determine the cross-linguistic differences in /θ/ production on the basis of the dissociation between perception and production of the English /θ/ (e.g., Brannen, 2002; Hancin-Bhatt, 1994a; Teasdale, 1997). Many phoneme identification and phoneme confusion studies have shown that /θ/ is perceptually most often confused with /f/ by native as well as by various nonnative listeners, and less frequently confused with /t/ or /s/ (e.g., Brannen, 2002; Cutler, Weber, Smits, & Cooper, 2004; Hancin-Bhatt, 1994a, 1994b; Miller & Nicely, 1955; Tabain, 1998). For example, Cutler et al. showed that under a 0-dB signal-to-noise ratio, Dutch L2 learners of English perceive English /θ/ 12.1% of the time as /θ/, 0.4% as /s/, 6.3% as /t/, and 13.3% as /f/

(for a comparison, American participants perceive /θ/ 18.3% of the time as /θ/, 0% as /s/, 5.4% as /t/, and 13.3% as /f/).<sup>1</sup> Hancin-Bhatt (1994a, 1994b) showed (using good listening conditions) that German L2 learners perceive English /θ/ 48% of the time as /θ/, 5% as /s/, 0% as /t/, and 22% as /f/. Despite methodological differences, both studies clearly show that /θ/ is perceptually most often confused with /f/ for both L2 and L1 listeners. Although it is well-attested that phoneme perception in L2 listening can be imprecise (e.g., Flege, 1995), the studies above show that Dutch and German listeners can distinguish between /t/, /s/, and /θ/ in English quite reliably, but experience some problems when distinguishing between /f/ and /θ/.

This general tendency to confuse /f/ and /θ/ could be due to the small phonetic distance between /f/ and /θ/, concerning mainly higher frequencies that are perceptually less salient (Jones, 2002). In addition, both fricatives share articulatory properties, being produced by forming a narrow constriction at the upper teeth (either with the lower lip, for /f/, or with the tongue, for /θ/). Given this similarity of /θ/ with /f/, it is rather surprising that among the most common production substitutes in English L2 are /t/ and /s/ (for an overview, see Brannen, 2002), even when /f/ is available in the L1 phoneme inventory of the L2 speakers. In addition to linguistic factors (see also Lombardi, 2003), non-linguistic variables (e.g., insufficient motivation, cultural habits) and formal language instruction have been shown to influence the accent-specific preference (e.g., Piske, MacKay, & Flege, 2001; for a review, see Flege 1988). Note that substitutions of /θ/ are not restricted to L2 speech; they also occur in dialects of English, with reported instances of /f/ in Cockney (Wells, 1982) and of /t/ in different dialects of English (Hickey, 2004; McGuire, 2003).

The most interesting aspect of /θ/ substitutions for the present study, however, is that the interdental fricative can typically be substituted in more than one way in a specific foreign accent. The different options vary in how frequently they occur and in their degrees of perceptual similarity to the canonical form. For example, while /s/ is clearly the predominant substitution choice for German learners of English (Hancin-Bhatt, 1994a), we also found in our production study (see further below) instances of /t/ and /f/. For Dutch listeners, in contrast, /t/ is the predominant substitution choice, but /s/ and /f/ substitutes have also been reported (Westers, Gilbers, & Lowie, 2007; see also our production study below). Note that /t/, /s/, and /f/ are all available in the L1 phoneme inventories of these L2 speakers. Thus, several variant forms can occur within a foreign accent and can

exhibit clear cross-linguistic differences in their frequencies of occurrence.

### Aims of the study

In the present study, we wanted to test how perceptual similarity and linguistic experience with L2 variants influence nonnative word recognition. If experience with L2 variants influences spoken-word recognition noticeably, we should find that frequent variant forms are recognized more easily than infrequent ones. However, if perceptual similarity is the driving factor in ease of L2 variant recognition, more similar variants should be recognized more easily than less similar ones. A natural group of listeners with experience with L2 variants are listeners who share a language background and a particular foreign accent. Thus, German L2 listeners have the most experience with the variant forms typical of a German accent in English (either from their own speech or from listening to other German learners of English), and Dutch L2 listeners have the most experience with the variant forms typical of a Dutch accent.

Using the eyetracking paradigm, we tested Dutch and German listeners' ability to recognize three /θ/ substitution types—/t/, /s/, and /f/—of English *th* words produced by either a Dutch or a German speaker. Eyetracking allows us to determine locations and latencies of fixations to pictures or printed words on a computer screen as the speech stream unfolds. Fixations are assumed to reflect the processes of lexical activation of a given word (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; for details on using printed words, see McQueen & Viebahn, 2007; Weber, Melinger, & Lara Tapia, 2007). The advantages of eye fixation over reaction time experiments are that the former experiments provide information about ongoing comprehension processes and that it is possible to arrange a design in which no task or explicit decision is required from the participants (see Huettig & Altmann, 2005).

In the present study, participants looked at a display of four printed words (e.g., *theft*, *left*, *kiss*, and *mask*) while they listened to English phrases containing one of the three substitution types (e.g., *theft* pronounced, with a /t/ substitution, as /tɛft/). Three different predictions were made: (1) Because /s/, /t/, and /f/ occur systematically in both the Dutch and German accents, the displayed *th* word *theft* would always be preferred as a potential target for the auditory input, as compared to a competitor word that also mismatched with the auditory input in the first segment, but for which the mismatch did not systematically occur as part of the listeners' accent (e.g., /tɛft/ does not systematically occur as a mispronunciation of the displayed competitor word *left*). (2) The frequencies of variants in L2 speech would modulate the degree of lexical activation. In this case,

<sup>1</sup> These percentages do not add up to 100% because we list only the confusion patterns involving /t/, /s/, and /f/ that are relevant for this study, rather than the entire confusion matrix.

Dutch listeners should preferably fixate *th* words when hearing /t/ substitutes, whereas German listeners should preferably fixate *th* when hearing /s/ substitutes, because these substitutes are the most frequent ones in the respective L2 speech. (3) Because /f/ is perceptually most similar to /θ/, /f/ substitutes (e.g., /fi:m/ for English *theme*) could prove a better match for *th* words than the accent-specific predominant substitutions /s/ and /t/, even though /f/ substitutions occur infrequently in German- and Dutch-accented English. Importantly, activation strength for /f/ substitutions should be comparable for the two listener groups, since /f/ is more confusable with /θ/ than is /s/ or /t/ for both Dutch and German listeners (Cutler et al., 2004; Hancin-Bhatt, 1994a, b).

By presenting half of the participants with recordings from a Dutch speaker and the other half with recordings from a German speaker, we wanted to control for the likely influence of L2-specific fine phonetic detail. For German listeners, substitutions produced by a Dutch speaker might prove a less good match for *th* words than were substitutions produced by a German speaker, and the reverse might be true for Dutch listeners. In this case, we should find overall stronger lexical activation of target words (e.g., *theft*) when the listener and speaker share a language background than when they differ in language background, and possibly also more pronounced effects of experience when the listener and speaker share a language background.

Although some empirical evidence for the /θ/ substitution preferences of German and Dutch speakers of English exists (see above), the evidence comes from studies that have used different materials and elicited the substitutions in different ways. Similarly, evidence for perceptual confusions involving /θ/ stems from studies that differed in their methodologies (see above). To have a controlled basis for examining the consequences of /θ/ substitutions for word recognition, it was important to establish perceptual confusability and substitution preferences for both Dutch and German speakers of English, using the same materials and elicitation methods.

For this purpose, we collected data of read speech with German and Dutch speakers and a control group of English speakers (just as for L2 speakers, the production data of the L1 speakers served as a control condition for the eyetracking study, as will be further discussed below). We also conducted an English AXB discrimination task with Dutch and German listeners. A discrimination task was administered because this task has often been used in previous investigations of the perception of L2 sound contrasts (e.g., Best, McRoberts, & Sithole, 1988; Best & Strange, 1992). Moreover, this task has the advantage of focusing on specific contrasts rather than testing a larger set of phonemes, as is

usually done in phoneme identification paradigms. The discrimination task will be presented first (Experiment 1a), to be followed by presentation of the production data (Experiment 1b). The eyetracking study will then be presented in Experiment 2.

## Experiment 1a: Discrimination study

### Method

#### Materials and design

In the perception task, the three English contrasts /θ/–/f/, /θ/–/s/, and /θ/–/t/ were tested. The materials consisted of four nonword syllables: /θ /, /f /, /s /, and /t /. A female native speaker of American English recorded each syllable 20 times (with a sampling rate of 44.1 kHz). For each of the syllables, five tokens were retained, so as to balance the syllable durations and *f*0 contours among the items to be contrasted. The retained tokens were used to make up AXB triplets with the four possible type orders AAB, ABB, BAA, and BBA, where A and B stand for the two members of a contrast. The inter-stimulus-interval (ISI) was 800 ms.<sup>2</sup> The X was always a different physical token than that of the categorically matched A or B.

#### Procedure

Each of the four order types was presented 10 times for each contrast, with a different combination of tokens for each AXB triplet. The 120 resulting trials (3 contrasts × 4 order types × 10 repetitions) were presented in randomized order to listeners over headphones. The listeners' task on each trial was to indicate (using a keyboard) whether the middle syllable X was the same as the first or the third syllable.

#### Participants

Groups of 14 native speakers of Dutch and 12 native speakers of German, all students from the subject pool of the Max Planck Institute for Psycholinguistics in Nijmegen, participated in the perception experiment.

<sup>2</sup> It could be claimed that an ISI of 800 ms might emphasize the comparison of abstract representations more than perceptual traces. We leave it to future studies to explore the effect of different ISIs on discrimination abilities.

## Results

The percentages of correct performance for the contrasts were analyzed in paired-samples *t* tests. For Dutch participants, the /θ/–/f/ contrast was the most difficult one (66.6% correct responses), with significantly worse performance than either the /θ/–/s/ contrast [88.6% correct responses;  $t(13) = 8.64, p < .001$ ] or the /θ/–/t/ contrast [90.5% correct;  $t(13) = 8.97, p < .001$ ]. The same pattern of results was observed for German participants, with performance being worse for the /θ/–/f/ contrast (73.5% correct responses) than for either the /θ/–/s/ contrast [88.5% correct responses;  $t(11) = 5.96, p < .001$ ] or the /θ/–/t/ contrast [90.0% correct responses;  $t(11) = 6.20, p < .001$ ]. The results of the AXB task thus confirm that /θ/ is indeed most confusable with /f/ for Dutch and German learners of English (see Cutler et al., 2004; Hancin-Bhatt, 1994a).

## Experiment 1b: Production study

### Method

#### Materials and design

A short story in English was used as the text, which consisted of 17 sentences, with a total of 18 tokens of word-initial /θ/ (these tokens consisted of 13 different content words occurring between 1 and 4 times in the text). Only three of these words also occurred in the eyetracking study. Since the production data were elicited immediately after participants had completed the eyetracking study, we deliberately avoided an overlap of words between the two studies in order to preclude an influence of the just-heard variants on the speakers' production of the same words.

#### Procedure

The participants were asked to read the text aloud at a comfortable speaking rate. The recordings were made in a quiet room with a portable digital recorder at a 44.1-kHz sampling rate with 16-bit resolution, and were later transferred to a computer. Using the PRAAT speech editor (Boersma, 2001), the initial consonant of the 18 tokens of words with initial voiceless /θ/ were labeled and then categorized by two trained research assistants. Whenever there was disagreement about the category of a particular token in the German and Dutch recordings (for the German data, the percentage of overall agreement was 79% and the kappa interrater reliability was .58; for the Dutch data, these numbers were 82% and .64, respectively), the categorization of a third labeler, a trained phonetician, was decisive. For the English recordings (the

percentage of overall agreement was 83%, and the kappa interrater reliability was .64), the labels of a native speaker were decisive. The raters used perceptual as well as acoustic–phonetic criteria (e.g., the presence of a closure, voicing, the amount of frication, the frequency range, and the periodicity). If the raters could not identify a sound on the basis of the acoustic–phonetic criteria, a perceptual judgment was made.

#### Participants

A total of 37 German learners of English (mean age 23) from the University of Cologne, 37 Dutch learners of English (mean age 21) from Radboud University in Nijmegen, and 31 native speakers of English (mean age 19) from the University of Birmingham participated in the production study. The L2 participants were highly proficient in English and had no difficulties reading and understanding the text. None of the German participants had lived in the Netherlands, and none of the Dutch participants had lived in Germany. However, Dutch students in Nijmegen (a city close to the German border) in general have more exposure to German than German students in Cologne have to Dutch.

### Results and discussion

The categorization results averaged over items showed that the German and Dutch speakers produced about half of their English *th* tokens as /θ/ (51% for the German speakers and 62% for the Dutch speakers). Within the substituted instances (see Table 1), a clear difference was found between German and Dutch speakers: Germans predominantly substituted the English fricative /θ/ with /s/ (23%) and less often with /t/ (5%), while Dutch speakers predominantly substituted /θ/ with /t/ (29%) and less often with /s/ (7%). For both L2 speaker groups, the perceptually similar /f/ was produced least frequently (3% and 5%, respectively).

It is worth mentioning that the choice of the substitution type (/t/, /s/, or /f/) was not necessarily speaker-specific. While some speakers consistently used one substitution, most speakers produced two or all three of the substitutions. Similarly, all three substitutions occurred for any given word, but with a very clear dominance of /s/ substitutions

**Table 1** Percentages of /θ/ substitutions per speaker group (percentages are rounded up, and numbers of occurrences are in parentheses)

Speaker Group	Substitution Type		
	F	S	T
German	3% (17)	23% (155)	5% (30)
Dutch	5% (34)	7% (43)	29% (187)
English	12% (63)	0% (0)	0.02% (1)

across all words for German speakers, and a clear dominance of /t/ substitutions across all words for Dutch speakers. Multiple /θ/ variants thus occurred within an accent and within a speaker.

The results for the English speakers showed that over 88% of *th* tokens were correctly pronounced with /θ/, and on average, 12% of the *th* tokens were pronounced with /f/ substitutions. A closer look at the results revealed that the majority of the /f/ substitutions were produced by three speakers, which was possibly a dialectal attribute (two of the speakers were from the area of Northampton and one speaker from the greater area of London). When excluding these three speakers from the analysis, the remaining 28 English speakers produced over 95% of the *th* words with /θ/.

Acoustic measures of the substitutions and of /θ/ productions were furthermore taken in order to evaluate possible differences in fine phonetic detail across the three speaker groups. In particular, duration, root-mean squared (RMS) amplitude, and center of gravity (COG) were measured (for further details, see Hanulíková & Weber, 2010). Planned *t* tests on the measures of the /θ/ productions showed that English L1 speakers differed from German L2 speakers in duration and RMS but not in COG, while Dutch L2 speakers differed from English L1 speakers in duration but not in RMS or COG. Differences in duration between L1 and L2 speakers are not surprising, given that L2 speech rates are slower overall. Similarly, differences in amplitude could come about when L2 speakers encounter difficulties with a given speech sound and consequently lower their voices in amplitude. A comparison of the three measurements for /θ/ realization between German and Dutch speakers did not show any significant differences. The English story used in the production study also contained tokens of intended /s/ and /t/ (10 words each) that were acoustically measured. In order to investigate whether substitutions are further marked with phonetic detail, accent-specific predominant substitutions (/t/ for Dutch and /s/ for German speakers) were compared with the realizations of intended instances of /t/ and /s/. Within the German speakers, /s/ substitutions differed from the intended /s/ productions as well as from the correctly pronounced /θ/ productions in all three measures. Similarly, Dutch speakers' /t/ substitutions differed from the intended /t/ productions in all three measures, and from /θ/ productions in duration and COG, but not in RMS. This suggests that although the substitutions in L2 speech are perceived and labelled as /t/ or /s/, they can differ in phonetic detail from those intended productions.

From the distribution of the substitutions, however, we can assume that Dutch and German listeners in the eyetracking study had indeed differing past experiences with /θ/ substitutions. If such past experience influences the recognition ease of these spoken variants, Dutch listeners should preferably

fixate *th* words when hearing /t/ substitutions, and German listeners should preferably fixate *th* words when hearing /s/ substitutions. This issue was examined in Experiment 2.

## Experiment 2: Eyetracking study

In what follows, we present a series of English eyetracking experiments with German and Dutch L2 listeners (Experiment 2a) and with a control group of English L1 listeners (Experiment 2b). The common methods and procedures for both experiments are followed by descriptions of the participants and the results, separated for the two experiments.

### Method

#### Materials and design

A total of 33 English content words with initial /θ/ were selected as target words. The target words were divided into three experimental groups (each group comprised 11 words), corresponding to the variants that listeners would hear: One group created substitution variants with initial /t/ (e.g., /tɛft/ for *theft*), a second group substitution variants with initial /s/ (e.g., /sɪf/ for *thief*), and a third group substitution variants with initial /f/ (e.g., /fi:m/ for *theme*). The variants did not yield other existing English words (except for the word *sunder*). Substitution Type was a between-items (but within-subjects) factor, because it was not possible to find enough target words for which each of the three substitutions resulted in a nonword.<sup>3</sup>

The log lemma frequencies per million of the *th* tokens were 1.20 in the /t/ condition, 1.61 in the /f/ condition, and 1.00 in the /s/ condition (determined with CELEX; Baayen, Piepenbrock, & Gulikers, 1995). There was no main effect of frequency in a univariate analysis of variance (ANOVA) [ $F(2, 30) = 1.85, p > .1$ ].<sup>4</sup> We also computed phonological neighborhood density using weighted density with the Kučera–Francis log frequencies of all neighbors (Vaden, Halpin, & Hickok, 2009). The density of *th* tokens was 14.78 in the /t/ condition, 12.56 in the /f/ condition, and

<sup>3</sup> Note that canonical forms of /θ/ were not added as an additional factor. Although a comparison with correct *th* productions could have informed us about the ease of recognition of canonical versus variant forms, we refrained from including correct *th* productions for two reasons. First, it would have been difficult to find L2 speakers with an audible accent that consistently produced /θ/ correctly (see our production study); and second, we know from previous research that mixing canonical and noncanonical pronunciations within an experiment can affect the processing of the noncanonical forms (Kraljic, Samuel, & Brennan, 2008).

<sup>4</sup> For more than two levels of comparison, we report Greenhouse–Geisser-corrected *p* values with uncorrected degrees of freedom.

5.87 in the /s/ condition. There was no main effect of density [ $F(2, 20) = 1.55, p > .2$ ].

Each target word was paired with a competitor that overlapped phonemically with the target to a large extent but differed in the initial consonant; for 29 of the 33 target words, the competitor and target in fact formed a minimal pair (e.g., competitor *left* for the target *theft*). Finally, for each target–competitor pair, two phonologically and semantically unrelated distractor words were selected (e.g., *kiss* and *mask* for the target–competitor pair *theft*–*left*). The log lemma frequency per million of each word in a quadruplet was balanced in such a way that the target word was on average not more frequent than the competitor or than any of the distractors. The averaged log frequencies for targets and competitors were 1.27 and 1.45, respectively, with no significant difference between them [ $F(1, 64) = 1.03, p > .1$ ]. The averaged log frequency per million of the two distractors was 1.53. The complete set of the experimental quadruplets is listed in the [Appendix](#).

During a trial, all four words of a quadruplet were shown on the screen as printed words. For the experimental trials, the target was presented auditorily with the respective /θ/ substitution, and therefore never fully matched any of the visually displayed words: Visually, the target words were presented in their correct spelling (e.g., *theft*), but auditorily, participants were presented with the substitution variants of targets (e.g., /tɛft/).

In addition to the 33 experimental quadruplets, 60 comparable filler quadruplets were created to prevent participants from developing the expectation that *th* words were the focus of the study. Thirty-five of the filler quadruplets included minimal pairs not containing *th* (e.g., *chain*, *gain*), which functioned either as filler target–competitor pairs or as distractors. When the minimal pairs functioned as distractors, words with initial /s/, /t/, or /f/ served as filler targets (which matched with the auditorily presented words), and words with initial /θ/ served as filler competitors (e.g., target *sane*, competitor *thirst*). In 25 of the filler quadruplets, none of the four words formed a minimal pair, and 16 of these 25 fillers contained visual words with *th*, but the auditorily presented word contained no /θ/. In 15 of these 25 fillers, the auditory input did not match with any of the displayed words at all (e.g., auditory /tɛnɪs/, visually *value*, *weather*, *noisy*, *singer*). The purpose of these fillers was to avoid having the participants expect all but the critical trials to have a fully matching word on the display, and thus to decrease the risk of any strategic behavior on the critical trials. In the remaining 10 fillers, the auditory input matched one of the displayed words (e.g., auditory /ɛnvi/, visually *envy*, *lower*, *image*, *brother*). An additional set of three practice quadruplets was constructed, none of which contained a visually presented word with *th*, and none of which contained *th*-pronunciation variants (see the [Appendix](#)).

All auditory target words were embedded in the English carrier sentence *Now you will hear. . .*. The complete set of sentences was recorded twice, once by a female native speaker of Dutch and once by a female native speaker of German. Both speakers were highly proficient in English but maintained an audible foreign accent. The speakers were chosen by the authors because they had an audible foreign accent. The strength of the accent was determined with ratings stemming from a questionnaire that all participants filled in after the eyetracking experiment. On a scale from 1 to 5 (with 1 meaning *no foreign accent* and 5 meaning *strong foreign accent*), L2 listeners rated the Dutch speaker on average with 3.48, and the German speaker with 3.46. Control ratings from L1 listeners were collected only for the German speaker, who obtained an average of 3.6. Both speakers were instructed to read the English sentences at a normal speech rate. They were not given any feedback on their pronunciation, but they were asked to produce the critical target words with /t/, /s/, or /f/ instead of /θ/.

Digital recordings were made in a soundproof booth at a 44.1-kHz sampling rate with 16-bit resolution and were directly saved onto a computer. The recordings were cut into single sentences, and the onset of the target words was labeled using PRAAT. Label points were determined auditorily and visually on the basis of oscillograms and spectrograms. The target word onsets were defined as follows: for voiceless /t/, by the onset of closure of the /t/, defined as the point where the vowel period of the preceding word has ceased, and for voiceless fricatives, by the onset of frication. All speech files were normalized so that their mean amplitudes were approximately equal. The speaking rate of the German speaker was faster than the speaking rate of the Dutch speaker, with the average duration of target words being 551 ms for the German speaker and 723 ms for the Dutch speaker [ $F(1, 32) = 261.32, p < .001$ ]. The average duration of the spoken context preceding the target was 736 ms for the German speaker and 829 ms for the Dutch speaker [ $F(1, 32) = 43.09, p < .001$ ]. For each participant, a randomized list with all critical items and all fillers was created, with the restriction that at least one filler occurred before an experimental trial. Participants were randomly assigned to one list with the recordings from either the German speaker or the Dutch speaker.

### Procedure

Participants were tested individually and were seated in front of a computer monitor. An SMI Eyelink II eyetracker (SR Research Ltd., Canada) was then calibrated for the dominant eye. Eye movements were monitored with a sampling rate of 500 Hz. The visual stimuli on each trial consisted of four printed words presented on the screen (1,024 × 768 resolution). Printed words were shown in black Times

New Roman in font size 34 against a white background. The four words on each trial were centered on four positions on the computer display ( $192 \times 256$  pixels,  $192 \times 768$  pixels,  $576 \times 256$  pixels, and  $576 \times 768$  pixels). The positions of the four words (target, competitor, and two distractors) were randomized across trials.

Participants received written instructions in English explaining that on each trial they would see four printed words on the screen, followed by an auditory sentence. They were told that on some trials, the word at the end of the auditory sentence would be displayed on the screen, but on other trials, the auditory word would not match with any of the printed words. Participants were instructed to look at the words on the screen while they listened to the sentences, but no explicit action was required (cf. Huettig & Altmann, 2005). A “look-and-listen” task rather than a “direct-action” task was used in order to avoid having participants make a decision in cases where variant forms did not match with any of the orthographically presented words. After the practice session, each participant was assigned to a new randomization list. At the start of each trial, a fixation cross appeared for 1,000 ms on the screen and was followed by a display with four printed words. The auditory sentences were played over headphones starting 800 ms after onset of the visual display for the recordings from the Dutch speaker, and starting 900 ms after onset of the visual display for the recordings from the German speaker. The increase in preview time for the German recordings was necessary to keep the timing between visual stimulus onset and target word onset comparable between the two experiments (due to the faster speaking rate of the German speaker). Note that the duration of the preview time follows previous research that has shown that longer preview times (e.g., about 1,000–1,500 ms) maximize phonological effects in the mapping process between a printed and a spoken word (Salverda & Tanenhaus, 2010).

The next trial started automatically 2,500 ms after the onset of the auditory sentence. A small fixation circle appeared on the screen after every six trials to initiate an automatic drift correction in the calibration of the eyetracker.

After completing the eyetracking experiment, participants were asked to read aloud the English text described in the production study. Participants then filled in a multiple-choice vocabulary test and a questionnaire about their language background, as well as ratings of the foreign accent in the auditory sentences of the eyetracking experiment. In the vocabulary test, 34 English words were presented with three possible translations given in the participants’ native language. Incorrect translations were either phonologically or semantically related to the critical word (e.g., for the word *thrower* in the German test, there were the three options: *Züchter* “breeder,” *Werfer* “thrower,” and *Schublade* “drawer”). Participants had to indicate which

of the three possibilities was the correct translation or to suggest another translation.

### Coding and analysis

For the analysis of the eyetracking experiments, the data from each participant’s dominant eye were used to determine the coordinates and timing of fixations. Only fixations that fell within a cell of one of the four interest areas—*target*, *competitor*, and two *distractors* (with a cell size of  $472 \times 344$  pixels, and a distance of 40 pixels between vertical cells and 60 pixels between horizontal cells)—were analyzed. Saccades were not added to fixation times.

The fixation proportions for the four interest areas were then calculated by aggregating all trials over participants and items in 20-ms steps in a time window from 0 to 1,000 ms after target word onset. The proportions were then converted to log odds (cf. Barr, 2008). An *odd* is a ratio of fixations to the critical region (e.g., the target) to fixations to all other regions (e.g., the competitor and the two distractors). We used odds rather than fixation proportions because proportions are bound by the values 0 and 1, while odds have the advantage of having an unbounded range, which is one of the assumptions for an ANOVA. A log odd of zero meant that the fixations to the critical region were equally likely to occur as not to occur (corresponding to a probability of .5). A positive log odd value meant that fixations to the critical region were likely to occur with a probability higher than .5, while negative values suggested that fixations to the target were likely to occur with a probability lower than .5.<sup>5</sup>

Statistical analyses were conducted over averaged values of empirical log odds (see, e.g., Agresti, 2002; Barr, 2008) for a time window of 200 to 900 ms after target word onset. The onset of 200 ms was chosen because it typically takes about 180 to 200 ms before a programmed eye movement is launched (see, e.g., Altmann & Kamide, 2004; Matin, Shao, & Boff, 1993). The offset of 900 ms was selected because it represented approximately the averaged target duration to which 200 ms were added, as well as representing the maximal duration of the target words. The main reason for an analysis window that represented the processing of the complete word rather than just the processing of the mismatch between variant forms and canonical forms was that biasing factors, such as lexical frequency effects and sublexical phonetic detail, do not necessarily dissipate as soon as discriminating information in the speech signal is encountered (e.g., Dahan, Magnuson, & Tanenhaus, 2001;

<sup>5</sup> For readers who are more familiar with proportions, fixation proportions to the target and competitor, and the mean of the distractor proportions are provided in a supplement to be found on the following homepage: <https://sites.google.com/site/adrihanulik/home/supplement>.



Salverda, Dahan, & McQueen, 2003). All inferential statistics in the eyetracking experiment were based on either  $t$  values or  $\min F'$  values.

### Experiment 2a: Dutch and German listeners

**Dutch participants** A total of 75 native speakers of Dutch, all students at Radboud University in Nijmegen, received payment for their participation. They had on average 7.4 years of formal training in English as a second language. In the multiple-choice vocabulary test conducted after the eyetracking experiment, they scored on average of 81.5% correct. A total of 37 students listened to recordings from the German speaker, and the other 38 students listened to recordings from the Dutch speaker.

**German participants** A total of 75 native speakers of German, all students at the University of Cologne, received payment for their participation. They had on average 8.7 years of formal English training. In the multiple-choice vocabulary test, the Germans scored on average 81.8% correct. A total of 37 of the German students listened to recordings from the Dutch speaker, and the other 38 students listened to recordings from the German speaker.

All participants reported normal hearing and normal or corrected-to-normal vision, and none reported any neurological impairment. The majority of the participants were trilingual, with French, German (for Dutch participants), or Spanish being the second nonnative language. For all participants, however, English was the first nonnative language.

### Results and discussion

Three of the Dutch and three of the German participants were excluded from further analyses because they did not follow the instructions of looking around the screen (i.e., they fixated the middle of the screen instead of the words). Fixations outside of the predefined regions of interest (see the [Coding and Analysis](#) section) were excluded from further analysis (on average, 18% of the data for the Dutch and German participants were excluded, with an even distribution across conditions).<sup>6</sup> Table 2 lists the averaged empirical log odds to the regions of interest from 200 to 900 ms after target word onset, averaged over the recordings from the German and the Dutch speakers. As can immediately be seen in Table 2, targets were always fixated approximately twice as much as distractors and competitors. Note that

between 0 and 200 ms, when fixations did not yet reflect the processing of the target word, there were never significantly more fixations to the target than to the competitor or the two distractors [all  $ps > .1$ ].

Although the visually displayed target and competitor both mismatched the initial consonant in the auditory input (e.g., both the target *theft* and the competitor *left* mismatched the initial sound in the heard variant /tɛft/), Dutch and German participants clearly interpreted the auditory input more readily as mispronounced *th* targets than as mispronounced competitors. This was true for all three auditorily presented /t/, /s/, and /f/ variants. This preference for matching all three variants onto *th* words instead of competitor words most likely reflects a combination of past experience and perceptual similarity. Due to the phonological overlap in rhyme between the competitor and the auditory input, competitors nevertheless received more fixations overall than did distractors (see Table 2). These rhyme effects were never present at the onset of the target word, but always emerged later, during the processing of the target word offset (see also Allopenna et al., 1998).

Having established that all three variant forms were readily interpreted as *th* targets, we will now concentrate on differences in the ease of variant recognition across listener groups and across speakers. In a first step, mean empirical log odds for both participants ( $F_1$ ) and items ( $F_2$ ) between 200 and 900 ms were entered into a three-factorial ANOVA with Substitution Type (/t/, /s/, or /f/) as a within-subjects and between-items factor, Speaker (German or Dutch) as a between-subjects factor, and Listener Group (German and Dutch) as a between-subjects factor. There were no main effect of substitution type [ $F_1(2, 280) = 3.46, p < .05; F_2(2, 80) = 4.49, p < .05; \min F'(2, 288) = 1.95, p > .1$ ] and no main effects of speaker [ $F_1(1, 140) = 3.15, p = .078, F_2(1, 40) = 3.96, p = .054; \min F'(1, 140) = 1.75, p > .1$ ] or listener group [ $F_1(1, 140) = 1.19, p > .1; F_2(1, 40) = 1.30, p > .1$ ]. There was also no three-way interaction [ $F_1(2, 280) = 1.24, p > .1; F_2(2, 80) = 1.31, p > .1$ ], but there was a significant two-way interaction between substitution type and listener group [ $F_1(2, 280) = 18.16, p < .001; F_2(2, 80) = 21.73, p < .001; \min F'(2, 274) = 9.89, p < .001$ ], suggesting that the L2 listeners differed in their looking preferences for *th* words upon hearing the substitution variants. To identify the source of the interaction, posttests were performed (see Table 2) and are discussed further below. Since there was no significant interaction between substitution type and speaker (both  $F_s < 1$ ), it can be assumed that Speaker was not a factor that significantly changed the recognition of variant forms. Thus, whether /tɛft/ was produced by a Dutch speaker or a German speaker had no significant influence on how Dutch and German listeners matched the variant form onto the *th* word. Therefore, all subsequent analyses do not further consider the Speaker factor and are conducted for

<sup>6</sup> The exclusion of fixations outside of the regions of interest did not alter the data pattern.

**Table 2** Averaged event likelihoods in empirical log odds to targets, competitors, and distractors within each condition, averaged across listener groups from 200 to 900 ms (Note that larger negative numbers mean fewer looks to a given region.)

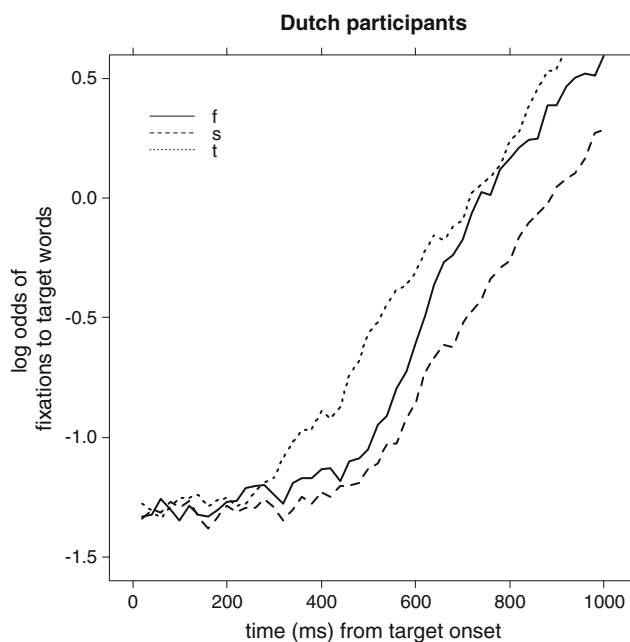
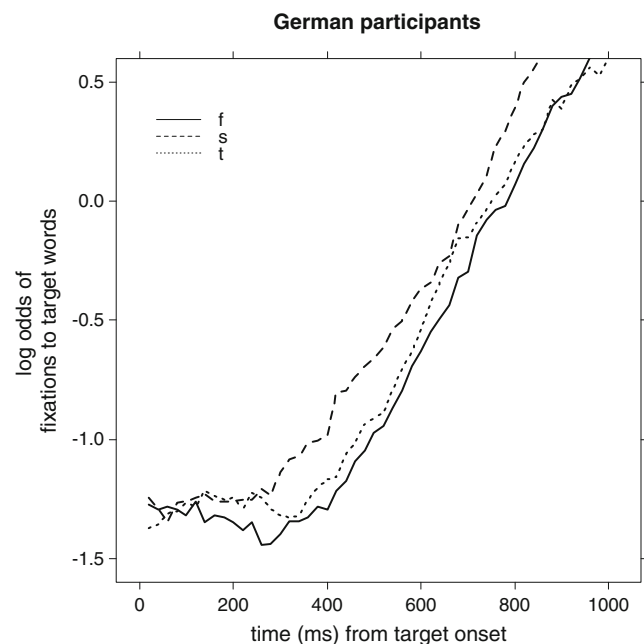
Listener Group	Substitution Type								
	/f/			/s/			/t/		
	Target	Competitor	Distractors	Target	Competitor	Distractors	Target	Competitor	Distractors
Dutch	-0.64	-1.66	-1.88	-0.85	-1.46	-1.72	-0.44	-1.61	-2.22
German	-0.69	-1.71	-1.89	-0.44	-1.63	-2.31	-0.63	-1.65	-1.99
English	-0.66	-1.77	-2.42	-0.86	-1.61	-2.49	-0.74	-1.72	-2.31

each listener group with data averaged across the German and Dutch versions of the experiment (as in Table 2).

Figures 1 and 2 show the fixation preferences for targets over the time from 0 to 1,000 ms after the acoustic onset of the /t/, /s/, and /f/-substituted target words, plotted separately for Dutch participants (Fig. 1) and German participants (Fig. 2). As can be seen in the figures, all three variants were recognizable as the intended target words (i.e., they ultimately directed participants' looks to the target), but there were clear differences between target fixations for the three substitution types in each listener group. This was confirmed by a main effect of substitution type for both Dutch participants [ $F_1(2, 140) = 16.97, p < .001$ ;  $F_2(2, 40) = 18.13, p < .001$ ;  $\min F^*(2, 129) = 8.77, p < .001$ ] and German participants [ $F_1(2, 140) = 5.81, p < .005$ ;  $F_2(2, 40) = 8.46, p < .002$ ;  $\min F^*(2, 150) = 3.44, p < .05$ ]. Note that these main effects are not spillover effects from early anticipatory fixations between 0 and 200 ms after target onset.

When fixations were not yet driven by acoustic input from the target word, there was no notable difference between target fixations when comparing the three substitution types (Dutch participants, both  $F_s < 1$ ; German participants, both  $F_s < 1$ ).

Planned  $t$  tests between substitution types within each listener group were then conducted (see Table 3). As predicted, the pairwise comparisons for Dutch participants showed that the /t/-substituted variants led to significantly more fixations to targets than did the less frequent /s/ variants or the perceptually similar /f/ variants. German participants, on the other hand, fixated targets significantly more after hearing the /s/ variants as compared to either the /t/ variants or /f/ variants. This suggests that the mapping process between auditory variants and the visual targets is driven more by linguistic experience than by perceptual similarity. Moreover, the time courses of the fixations further attest to the earliness of the effects. As can be seen in

**Fig. 1** Results of the eyetracking study with Dutch listeners, averaged over Dutch and German speakers: Log odds of fixations to target words after hearing variants with /t/, /s/, or /f/ substituted for /θ/**Fig. 2** Results of the eyetracking study with German listeners, averaged over Dutch and German speakers: Log odds of fixations to target words after hearing variants with /t/, /s/, or /f/ substituted for /θ/

Figs. 1 and 2, fixations to the target started to rise earliest for the dominant variant within each listener group. Thus, Dutch participants fixated the /t/ variant earlier than the other two variants. Similarly, German participants fixated the /s/ variant earlier than the other two variants. Both effects started to emerge between 200 and 400 ms after target onset, and were significant as early as 400–600 ms after word onset (all  $ps < .01$ ). The interpretation of *th* substitutions thus occurs immediately, well before the complete word is heard.

Taken together, Experiment 2a showed that fixation preferences match with linguistic experience as determined by the accent-specific frequencies of variants. The most frequently produced variant in a respective L2 group was the most easily recognized form. Although /f/ is perceptually most similar to /θ/, *th* words were not activated more strongly when they were pronounced with a /f/ substitute than when they were pronounced with the accent-specific predominant substitute. This suggests that, at least in the case of *th* substitutions, the influence of perceptual similarity does not outweigh that of linguistic experience with respect to the phonetic-to-lexical mapping process.

The role of perceptual similarity in L2 variant recognition, however, cannot be completely neglected when comparing the ease of recognition of the two less frequent variants within each speaker group. While the production study has shown that both German and Dutch learners produced /f/ variants slightly less often than /t/ variants or /s/ variants, the fixation probabilities did not seem to exactly mirror this dispreference for /f/ forms. That is, Dutch speakers produced 5% /f/ variants and 7% /s/ variants, but in the eyetracking study they recognized /f/ substitutions faster than /s/ substitutions. German participants produced 3% /f/ variants and 5% /t/ variants, but they recognized /f/ substitutions as fast as /t/ substitutions. This suggests that within the infrequent substitutes, the perceptually similar variant outperforms or is as good as the less similar variant.

Apparently, for neither of the listener groups was the influence of experience on lexical activation modulated by different speakers. Independently of whether Dutch

listeners heard a German or a Dutch speaker, they preferably fixated targets after /t/ variants. Similarly, German participants preferably fixated targets after /s/ variants, independently of the speaker. This suggests that L2 lexical processing of frequent variant forms is indeed influenced by linguistic experience with the variants from the listeners' own accent, and is furthermore robust enough not to be disrupted by fine phonetic differences in the production of *th* variants by speakers with a different language background.

Although the results from Experiment 2a seem clear cut, to fully evaluate the contribution of perceptual similarity to variant recognition requires a control group of native listeners of English. In Experiment 2b, we therefore compared the performance of English L1 listeners to that of the L2 listeners. The English speakers had probably not previously experienced the three substitutions in any way similar to the German and Dutch learners. However, it is reasonable to assume that English listeners have some experience with *th* substitutions. This may stem both from speaking and listening to dialects of English (e.g., /f/ and /t/ variants; see McGuire, 2003; Wells, 1982) and from listening to a variety of L2 speakers of English. We hypothesized that the perceptually similar /f/ variant should lead to a stronger activation of *th* words, if any at all. This was so for two reasons: First, the effect of experience with /θ/ substitutions was less pronounced in the production of the English listeners in the present study, but there were few instances of /f/ substitutions. And second, for English listeners, just as for L2 listeners, /f/ is perceptually most similar to /θ/ (e.g., Cutler et al., 2004).

The procedure of Experiment 2b was identical to that of Experiment 2a, except that English listeners in Experiment 2b only heard the recordings from the German L2 speaker. Since we did not find any interaction between substitution type and speaker in Experiment 2a, we assumed that the language background of the L2 speaker would not play a significant role in the strength of word activation for English listeners in Experiment 2b, either.

**Experiment 2b: English listeners**

*Participants* A total of 34 native speakers of English, all students at the University of Birmingham, received payment for their participation. All of the participants reported normal hearing and normal or corrected-to-normal vision, and none reported any neurological impairment.

**Results and discussion**

Three of the participants were excluded because of technical problems or because they did not perform the task of

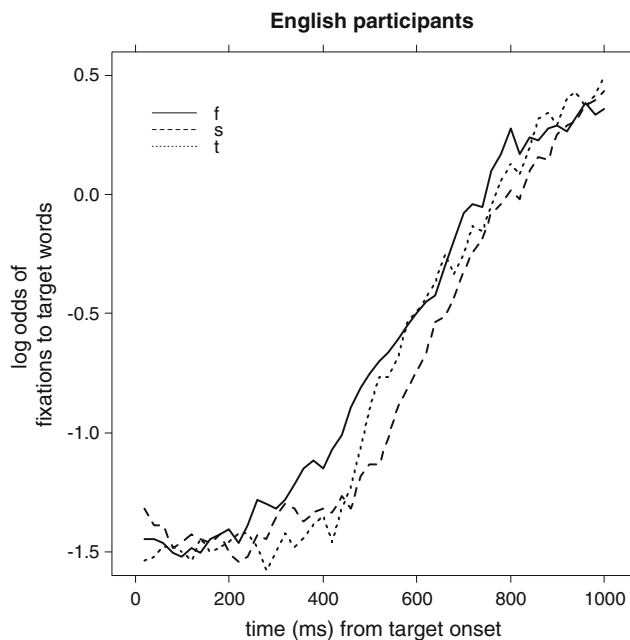
**Table 3** Paired *t* tests

Substitution	Dutch Participants	German Participants	English Participants
/f/ vs. /s/	$t_1(71)=3.06^{**}$	$t_1(71)=-3.00^{**}$	$t_1(30)=1.86$
	$t_2(20)=3.84^{**}$	$t_2(20)=-3.74^{**}$	$t_2(20)=1.35$
/s/ vs. /t/	$t_1(71)=-5.83^{**}$	$t_1(71)=2.68^{**}$	$t_1(30)=0.86$
	$t_2(20)=-5.22^{**}$	$t_2(20)=2.76^*$	$t_2(20)=0.63$
/f/ vs. /t/	$t_1(71)=-2.74^{**}$	$t_1(71)=-0.74$	$t_1(30)=1.04$
	$t_2(20)=-2.79^*$	$t_2(20)=-1.43$	$t_2(20)=0.97$

$^{**}p < .01$ ,  $^*$ Bonferroni corrected  $p < .02$

looking around on the screen. Fixations outside of the predefined regions of interest were again excluded from further analysis (on average, 19% of the data, with an even distribution across conditions). Table 2 lists the averaged empirical log odds to targets, competitors, and the averaged distractors. As can be seen in Table 2, English listeners fixated targets at least twice as often as any of the distractors or competitors (with no significant differences in the time window 0–200 ms), suggesting that English listeners, just like Dutch and German listeners in Experiment 2a, interpreted all three variant forms as *th* words. Figure 3 shows looking preferences of the English listeners for *th* target words when hearing the /t/, /s/, and /f/ variants, averaged over all participants in 20-ms time steps. Visual inspection of Fig. 3 reveals that all three variants were interpreted as the intended targets. However, in contrast to Experiment 2a, ANOVAs showed no main effect of substitution type [ $F_1(2, 60) = 1.39, p = .26; F_2 < 1$ ].

Given the comparable increases of fixations to target words for all three substitution variants in Experiment 2b, we can assume that English participants activated the underlying representations of the intended *th* words with equal strength. However, the time courses of the fixations in Fig. 3 suggest that fixations to the target were earliest for the perceptually similar /f/ variants. Even though this effect is visible in a 200- to 400-ms time window, it was not significant. A marginally significant difference was only found for a comparison of /f/ substitutions to /s/ (but not to /t/ substitutions) as early as 400–600 ms after word onset



**Fig. 3** Results of the eyetracking study with U.K. English participants listening to a German speaker: Log odds of fixations to targets words after hearing variants with /t/, /s/, or /f/ substituted for /θ/

[significant across participants only:  $t_1(30) = 2.05, p = .05$ , an uncorrected  $p$  value].

This result does not confirm previous findings that perceptually similar forms disrupt L1 word recognition less strongly than perceptually dissimilar forms (e.g., Connine et al., 1993; Connine et al., 1997; Marslen-Wilson & Zwitserlood, 1989). However, it could also be that the role of perceptual similarity was not pronounced due to familiarity with all three substitution types. After all, the substitution of the interdental fricative is a well-known phenomenon, not only within nonnative communities, but also across native speakers of English. English participants very likely had past encounters with all three variant forms, either from listening to English speakers with a regional dialect or from listening to a variety of L2 speakers. It is therefore possible that the amounts of experience with /s/ and /t/ variants were sufficiently high that, even for L1 listeners, the perceptual confusability of /f/ variants did not outweigh the other two variants as strongly as expected.

## General discussion

This study investigated the production and recognition of foreign-accented variants of English words beginning with the voiceless fricative /θ/. A production study confirmed earlier results that both German and Dutch learners of English exhibit differences in the production frequencies of /t/, /s/, and /f/ substitutions for /θ/, despite comparable discrimination difficulties in perception (as confirmed in the AXB task). While Dutch learners predominantly substituted /θ/ with /t/, German learners substituted /θ/ most often with /s/. The perceptually most similar substitute /f/ occurred least often within the two groups. In contrast to L2 learners, L1 speakers showed only a few /θ/ substitutions, all of which were /f/ substitutions.

Using eyetracking, we then showed that both L2 and L1 listeners can successfully compensate for pronunciation variation in foreign-accented speech. Both Dutch and German L2 listeners, as well as a control group of English L1 listeners, interpreted the intended *th* words when hearing all three variants: /t/, /s/, and /f/. This was revealed by a greater number of fixations to *th* words than to competitor words (e.g., when hearing /tεft/, listeners fixated *theft* and not *left*). This result suggests that our listeners have had experience with or knowledge about possible /θ/ variants in English. This experience was apparently sufficient to allow mapping of the variants to *th* words rather than to other words with an initial mismatch.

Interestingly, however, the ease of mapping of the variants to *th* words was modulated by listeners' background. Dutch listeners activated English *th* words more strongly

when they were pronounced with a /t/ substitute than when they were pronounced with a /s/ substitute. In contrast, German listeners activated *th* words more strongly when they were pronounced with a /s/ substitute than when they were pronounced with a /t/ substitute. L1 listeners did not show any clear preference in variant recognition. Thus, experience from either producing or listening to one's own accent influenced word recognition in L2, such that accent-specific predominant variants were recognized most easily. Accent-specific infrequent variants (e.g., /t/ for German speakers and /s/ for Dutch speakers) were recognized either as fast as, or less easily than, the least frequent, but perceptually most similar, /f/ variant. This suggests that experiential effects in variant recognition can outweigh the effects of perceptual similarity. Note that even though we cannot determine how much of the general ease of recognition is influenced by the closed set of words used in eyetracking designs, studies using open-ended tasks such as cross-modal priming have shown that the recognition of comparable variants is possible, as well (Weber, Broersma, & Aoyagi, 2011; Weber, Sumner, Krott, Huettig, & Hanulíková, 2011). The disadvantages of open-ended tasks are, however, that they require an explicit decision and do not provide information about the time course of processing.

Perhaps quite surprising was the finding that the pattern of results for L2 listeners in the present study held irrespective of the language background of the L2 speaker (English spoken by a German or a Dutch learner), pointing to general and not to accent-specific variant recognition. It is possible, however, that the English productions of the German and Dutch speakers were simply more similar than expected. Even though we know that /s/ and /t/ differ phonetically in Dutch and German (Lisker & Abramson, 1964; Mees & Collins, 1982), it is possible that—when speaking English—our speakers' variants of /s/ and /t/ were acoustically and perceptually less different from each other. To address this possibility, acoustic analyses of the recorded material were conducted on duration and COG. Both the /s/ and /t/ substitutions spoken by the German speaker differed in COG from those of the Dutch speaker [both  $ps \leq .05$ ], and only /s/ also differed in duration (the German /s/ was longer than the Dutch /s/) [ $t(10) = 2.27$ ,  $p < .05$ ]. Thus, there were indeed differences in the realizations of the English /s/ and /t/, but these differences apparently did not impact listeners' variant recognition. On the other hand, it could also be that through prior exposure to multiple speakers, our listeners have generalized their experience across L2 speakers, even across L2 speakers with different L1 language backgrounds. This outcome is in line with previous studies on accented-speech processing, in general, and on short-term perceptual learning, in particular, according to which generalization processes can apply to multiple speakers and multiple

accents (see, e.g., Bradlow & Bent, 2008; Sidaras, Alexander, & Nygaard, 2009).

The difference in ease of recognition of /θ/ variants between L1 and L2 listeners leads to the question of which type of prior experience modulates variant recognition. For instance, Bell-Berti, Raphael, Pisoni, and Sawusch (1979) proposed that speech perception is influenced by how listeners, as speakers, produce speech. On the other hand, Massaro (1987) and Diehl and Kluender (1989) argued that the past experience that listeners have with hearing other people speak creates memories of acoustic cues for phonemes or syllables that are associated with their underlying mental concepts. But how does the present study contribute to the debate about the roles of production and perception frequency in word recognition? Is the recognition of variant forms mainly driven by the L2 speakers' own productions, or is listening to other speakers with the same accent more important? It is of course impossible to exactly determine how much of our listeners' experience stemmed from producing the variant forms themselves and how much stemmed from listening to other L2 speakers (after all, even the most talented learners of English will have gone through an initial learning phase in which they produced some /θ/ variants). However, in an attempt to assess the role of production in looking preferences, we carried out correlation analyses to test whether individual production frequencies from our production study predicted looking behavior in the eyetracking study. For this purpose, we compared the differences in *th*-word fixations to the differences in production frequencies for each participant. For each of the three substitution types (/t/, /s/, and /f/), we calculated the fixation differences (e.g., /t/-/s/, /s/-/f/, and /f/-/t/), and compared those to the analogous differences in production frequency. We did not find any significant correlation between looking preferences and production data for L2 listeners. Note that an obvious weakness of this test is the use of different words in the production study and the eyetracking study, since productions of substitutions might vary somewhat on the level of individual words. The results suggest that looking preferences may not be driven by individual production frequencies, but it should be considered that this conclusion is based on a null result.

We also learned from the present study that the mapping from the variant forms to the lexical representations of the intended words happens fast. While most of the previous studies on variant recognition took measures after the complete variant form was heard (either using a lexical decision task or cross-modal priming), the eyetracking method in the present study allowed us to show that the compensation for the variation takes place while, and not only after, the complete word has been heard. This might suggest that a syllable beginning with, say, /s/ (or possibly the /s/ segment itself) by rule not only maps onto words starting with /s/ in

the lexicon, but also onto words starting with /θ/, before more evidence for an existing *th* word is available.

This is the first time that experiential effects on L2-variant recognition have been shown for variant forms that are perceptually conspicuous (i.e., variant forms that listeners are typically aware of).<sup>7</sup> Earlier studies had investigated the recognition of variant forms that result from phonological processes such as /t/ lenition or flapping in L1 speech (Connine, 2004; Mitterer & Ernestus, 2006; Ranbom & Connine, 2007). These processes are contextually driven sound changes that usually ease articulation; as a result, the segments surrounding the sound change, and the changed segment itself will overlap more with their phonological features (e.g., the final nasal of *lean* in *lean bacon* shares place of articulation with the following stop after assimilation). According to some phonological approaches, such phonological rules tend to lead to perceptually inconspicuous changes from the canonical form (e.g., Hura, Lindblom, & Diehl, 1992; Steriade, 2001). Although the frequency and the distribution of variant forms need to be considered in addition to perceptual changes (e.g., Ranbom et al., 2009), it can be said that the most common changes in L1 are not easily noticed. As a consequence, not only experience, but also basic auditory processes, can be instrumental in the compensation for these phonological processes (Mitterer, Csépe, Honbolygo, & Blomert, 2006). This is not the case for /θ/ substitutions in L2 speech, where the choice of phoneme substitution rarely coincides with the perceptually least noticeable option (i.e., the predominant Dutch and German substitutes /t/ and /s/ are perceptually distinct from English /θ/, and the perceptually similar /f/ is the least preferred substitute). The fact that experiential effects were also found for this type of variant suggests that compensation for segmental substitutions in foreign-accented speech goes beyond basic auditory processes.

This result raises the issue of how words and their variants are represented in the L2 lexicon. Current theories of L1 processing can be divided into representation-based and processing-based accounts. Some representation-based accounts postulate that pronunciation variation is encoded in lexical entries. Episodic models of this type maintain that every instance of a word ever encountered is encoded with fine phonetic detail in its corresponding mental representation (e.g., Goldinger, 1996; Johnson, 2006). Episodic traces of variant forms would thus exist alongside traces of canonical forms of a word. Ranbom and Connine (2007) also assumed that multiple forms are stored in our mental lexicon, but instead of multiple episodic traces, they suggested multiple abstract representations for variant forms. Thus, the flap variant *gentle*

for English *gentle* would have a separate abstract lexical representation. Yet another view is taken by Lahiri and colleagues (Lahiri & Marslen-Wilson, 1991; Lahiri & Reetz, 2002), who proposed that lexical entries are not only abstract but that they are also phonologically underspecified. In this model, a lexical entry for a given word would only include phonological features that are reliably associated with tokens of that word. Despite differences in the assumptions about the nature of lexical representations, these models have in common that listeners easily recognize variants because lexical representations reflect the variability encountered in the speech signal.

The core idea of processing-based accounts, on the other hand, is that listeners can compensate for variation by using segmental context. These accounts generally concentrate on variation that is created by continuous speech processes—that is, coarticulation. Again, different mechanisms have been proposed, ranging from the recovery of speech gestures from the signal (Fowler, 1996; Liberman, 1996), to general auditory mechanisms (Lotto & Holt, 2006; Mitterer et al., 2006), to more abstract pattern recognition mechanisms (Smits, 2001). An example of the latter is the phonological-inference account of Gaskell and Marslen-Wilson (1998), in which phonological rules rectify the variant forms into the canonical form. Gow (2002) has a somewhat less abstract view, according to which principles of perceptual organization help to compensate for pronunciation variation.

Although the present study was not set up to resolve the debate as to the nature of lexical representation or the question about the exact mechanisms underlying the compensation for variant forms, the results clearly suggest a role for linguistic experience in nonnative listeners. We can infer on the basis of the strengths of lexical activation for different phonological variants in the present study that the more-frequent variants activated the underlying lexical representations more strongly (as in Ranbom & Connine, 2007). According to Ranbom and Connine, listeners' lexical representations are influenced by the statistical properties of language. On their view, abstract phonological representations of frequent variants are stored in the mental lexicon along with the canonical forms. Less-frequent forms can also be stored, but their representations will be less strongly linked to the lexical entry than are the frequently produced forms. Thus, the stronger activation of predominantly produced variants could suggest that frequent variants are stored as phonological variants together with the canonical form. Stored lexical representations of variants may not even be so surprising for second language learners who learn a second language while being resident in their native country, where the experience with the L2 likely stems from talking to other nonnative speakers. It is even possible that L2 listeners would encounter the variant forms before the canonical forms. However, it is also possible that all three variants would be stored, given that all three showed stronger strength of activation as compared to arbitrary variants.

<sup>7</sup> Note that listeners did not comment after the experiment that they noticed a mismatch between the auditory and visual words. However, some participants did comment on the “bad” English pronunciation of the speakers.

An alternative proposal is that, on the basis of past experience, learners of English might have learned to map different /θ/ substitutions onto the underlying representations by rule, applying an abstract phonetic-to-lexical mapping. In such a case, they would activate all /s/-, /t/-, and /f/-initial words, as well as /θ/-initial words, whenever they encounter /s/, /t/, and /f/ in the speech input. This could point to a generalization mechanism or mediated access in word recognition, as suggested by models such as Shortlist (Norris, 1994). Phonemic variation would then map onto more-abstract underlying phonological categories. Thus, if learners were presented with new words not previously heard, they should generalize the abstract phonological knowledge to this new input. Future research could further investigate this issue using pseudo-words. Although we cannot clearly determine which of the existing models is correct, on the basis of the present results, it seems that recognition of variant forms is based on generalization of abstract phonological knowledge to the speech input. This generalization seems to be modulated by the frequency of substitutions occurring in foreign-accented speech. This would be in line with current models that assume abstract lexical representations for pronunciation variants of a word, which are informed by prior exposure to productions of that word.

An alternative account of the present results could be based on perceptual learning of phoneme categories. This account would place the effect at a phoneme level at which the L2 speakers had broadened their category of /θ/ to include the phonemes found in the alternative variants, perhaps most efficiently so when the substitution is a frequent one. One way to possibly examine this issue in future research would be to see whether the existing phoneme categories in the L1 are altered on the basis of experience with the variant substitutions in L2-accented speech.

To conclude, the present study suggests that if a word spoken in one’s L2 varies in how often it is mispronounced by L2 speakers, nonnative listeners make use of the frequencies of the pronunciation variants when recognizing L2 words in foreign-accented speech. They do so independently of whether they listen to L2 speakers with whom they share their L1 or to speakers with a different L1. Although perceptual similarity also modulates L2-variant recognition, the primary influence stems from learners’ experience with frequent variants. German travelers on board a German train hence should easily recognize the intended word *thank* behind the substitution variant *senk* in the conductor’s announcement *Ve vant to senk all pessengers for trevelling viz Deutsche Bahn*, but Dutch and English travelers should need just a bit longer to understand the intended word.

**Author note** We thank Katja Suckow for her assistance in this project, Debra Malpass for running the experiment in Birmingham, and Dale Barr for statistical advice. For further assistance in the

production study, we thank Laurence Bruggeman, Anne Blankenhorn, Sabrina Jung, Julia Lennertz, Simon Mack, Berit Meinert, Rachel Sheer, and Karina Visser. For constructive feedback on the manuscript, we thank Frank Eisner, Mark Pitt, Cynthia Connine, and two anonymous reviewers. This research was funded by the Max-Planck-Gesellschaft, Germany.

### Appendix

**Table 4** Stimulus materials

Spoken Word	Target	Competitor	Distractors	
campus	campus (P)	cigar	career	access
racket	racket (P)	supper	matter	polish
chap	sheep (P)	keep	nose	shave
teft	theft	left	kiss	mask
tings	things	kings	post	break
torough	thorough	borough	sickness	exam
taw	thaw	raw	key	noon
tew	thew	pew	kite	mint
ting	thing	ring	play	change
tinner	thinner	winner	pocket	heavy
tud	thud	mud	scar	bike
tump	thump	pump	sue	globe
troat	throat	growth	pure	harsh
trifty	thrifty	shifty	packing	flake
srive	thrive	drive	finish	basic
sief	thief	chief	fed	beard
srash	thrash	crash	freak	blade
sread	thread	bread	freeze	cube
srust	thrust	crust	keen	walk
srill	thrill	drill	fund	cave
srice	thrice	price	kid	blood
srower	thrower	grower	kettle	lemon
srough	through	brew	frame	hall
sunder	thunder	wonder	kindly	pillow
sermal	thermal	kernel	falter	willow
fanks	thanks	ranks	sick	nail
fimble	thimble	nimble	penny	booklet
forn	thorn	corn	proud	bench
frew	threw	crew	sole	clean
frow	throw	grow	press	bitter
fumb	thumb	dumb	song	glue
fank	thank	bank	park	rub
firty	thirty	dirty	stupid	artist
fird	third	bird	peel	wave
feme	theme	deem	print	awe
firteen	thirteen	sixteen	butcher	humid

P stands for a practice word.

## References

- Agresti, A. (2002). *Categorical data analysis*. Hoboken, NJ: Wiley.
- Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439. doi:10.1006/jmla.1997.2558
- Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. Henderson & F. Ferreira (Eds.), *The integration of language, vision and action* (pp. 347–386). New York, NY: Psychology Press.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457–474.
- Bell-Berti, F., Raphael, L. J., Pisoni, D. B., & Sawusch, J. R. (1979). Some relationships between speech production and perception. *Phonetica*, 36, 373–383.
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345–360. doi:10.1037/0096-1523.14.3.345
- Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20, 305–330.
- Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glott International*, 5, 341–345.
- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106, 707–729.
- Brannen, K. (2002). The role of perception in differential substitution. *Canadian Journal of Linguistics*, 47, 1–46.
- Coenen, E., Zwitserlood, P., & Bólte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language & Cognitive Processes*, 16, 535–564.
- Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review*, 11, 1084–1089. doi:10.3758/BF03196741
- Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32, 193–210.
- Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics*, 70, 403–411. doi:10.3758/PP.70.3.403
- Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, 37, 463–480.
- Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America*, 116, 3668–3678.
- Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, 42, 317–367.
- Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, 1, 121–144.
- Flege, J. E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders: A review—1988*. Norwood, NJ: Ablex.
- Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233–272). Baltimore, MD: York Press.
- Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, 99, 1730–1741.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 144–158. doi:10.1037/0096-1523.22.1.144
- Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 380–396. doi:10.1037/0096-1523.24.2.380
- Gaskell, M. G., & Snoeren, N. D. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1632–1647.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–1183.
- Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28, 163–179.
- Hancin-Bhatt, B. J. (1994a). *Phonological transfer in second language perception and production*. University of Illinois: Unpublished doctoral dissertation.
- Hancin-Bhatt, B. J. (1994b). Segment transfer: A consequence of a dynamic system. *Second Language Research*, 10, 241–269.
- Hanulíková, A., & Weber, A. (2010). Production of English interdental fricatives by Dutch, German, and English speakers. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *New sounds: Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech, Poznań, Poland* (pp. 173–178). Poznań: Adam Mickiewicz University.
- Hickey, R. (ed.) (2004). *A sound atlas of Irish English*. Berlin: Mouton de Gruyter.
- Huetting, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, 96, 23–32.
- Hura, S. L., Lindblom, B., & Diehl, R. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech*, 35, 59–72.
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34, 485–499.
- Jones, M. J. (2002). More on the “instability” of interdental fricatives: Gothic *þliuhan* “flee” and Old English *flēon* “flee” revisited. *Word*, 53, 1–8.
- Klatt, D. (1986). The problem of variability in speech recognition and in models of speech perception. In J. S. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 301–324). Hillsdale, NJ: Erlbaum.
- Kohler, K. J. (1991). The phonetics/phonology issue in the study of articulatory reduction. *Phonetica*, 48, 180–192. doi:10.1159/000261883
- Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, 19, 332–338.
- Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245–294.
- Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 637–675). Berlin: Mouton de Gruyter.



- Liberman, A. M. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops—Acoustical measurements. *Word*, *20*, 384–422.
- Lombardi, L. (2003). Second language data and constraints on manner: Explaining substitutions for the English interdentalals. *Second Language Research*, *19*, 225–250.
- Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics*, *68*, 178–183. doi:10.3758/BF03193667
- Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 576–585.
- Massaro, D. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.
- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*, 372–380. doi:10.3758/BF03206780
- McGuire, G. (2003, February). *The realization of interdental fricatives in Columbus, OH, AAVE*. Paper presented at the Montreal–Ottawa–Toronto Phonology Workshop. Retrieved February 2009 from [www.ling.ohio-state.edu/~mcguire/Interdentals/Handout.doc](http://www.ling.ohio-state.edu/~mcguire/Interdentals/Handout.doc)
- McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, *60*, 661–671.
- Mees, I., & Collins, B. (1982). A phonetic description of the consonant system of Standard Dutch. *Journal of the International Phonetic Association*, *12*, 2–12.
- Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, *27*, 338–352.
- Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science*, *30*, 451–479.
- Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, *34*, 73–103.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191–215.
- Pitt, M. A. (2009). The strength and time course of lexical activation of pronunciation variants. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 896–910. doi:10.1037/a0013160
- Pitt, M. A., Dilley, L. C., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics*, *39*, 304–311.
- Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*, 273–298. doi:10.1016/j.jml.2007.04.001
- Ranbom, L. J., Connine, C. M., & Yudman, E. M. (2009). Is phonological context always used to recognize variant forms in spoken word recognition? The role of variant frequency and context distribution. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1205–1220. doi:10.1037/a0015022
- Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89. doi:10.1016/S0010-0277(03)00139-2
- Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1108–1117. doi:10.1037/a0019901
- Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior*, *23*, 450–466.
- Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, *52*, 240–255. doi:10.1016/j.jml.2004.11.001
- Sebastián-Gallés, N., Vera-Constán, F., Larsson, J. P., Costa, A., & Deco, G. (2009). Lexical plasticity in early bilinguals does not alter phoneme categories: II. Experimental evidence. *Journal of Cognitive Neuroscience*, *21*, 2343–2357.
- Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America*, *125*, 3306–3316.
- Smits, R. (2001). Hierarchical categorization of coarticulated phonemes: A theoretical analysis. *Perception & Psychophysics*, *63*, 1109–1139.
- Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, *48*, 233–254.
- Steriade, D. (2001). Directional asymmetries in place assimilation: A perceptual account. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 258–275). New York, NY: Academic Press.
- Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*, 487–501.
- Tabain, M. (1998). Non-sibilant fricatives in English: Spectral information above 10 kHz. *Phonetica*, *55*, 107–130.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634. doi:10.1126/science.7777863
- Teasdale, A. (1997). On the differential substitution of English [θ]: A phonetic approach. *Calgary Working Papers in Linguistics*, *19*, 71–91.
- Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). *Irvine Phonotactic Online Dictionary, Version 2.0* [Data file]. Available at [www.iphod.com](http://www.iphod.com)
- Weber, A., Broersma, M., & Aoyagi, M. (2011a). Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics*, *39*, 479–491. doi:10.1016/j.wocn.2010.12.004
- Weber, A., Melinger, A., & Lara Tapia, L. (2007). The mapping of phonetic information to lexical presentations in Spanish: Evidence from eye movements In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 1941–1944). Dudweiler: Pirrot.
- Weber, A., Sumner, M., Krott, A., Huettig, F., & Hanulíková, A. (2011, June). *Sinking about boats and brains: Activation of word meaning in foreign-accented speech by native and nonnative listeners*. Poster presented at the First International Conference on Cognitive Hearing Science for Communication, Linköping, Sweden.
- Wells, J. C. (1982). *Accents of English 2: The British isles*. Cambridge, U.K.: Cambridge University Press.
- Westers, F., Gilbers, D., & Lowie, W. (2007). Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences*, *29*, 477–491.