



Increased exposure and phonetic context help listeners recognize allophonic variants

Malte C. Viebahn¹ · Paul A. Luce²

Published online: 24 April 2018
© The Psychonomic Society, Inc. 2018

Abstract

This study examines the influence of increased exposure and phonetic context on the recognition of words that are produced with nasal flaps in American English (e.g., the word *center* produced as *cenner*). Previous work has shown that despite their high frequency of occurrence, words produced with nasal flaps are recognized more slowly and less accurately compared with canonical pronunciation variants produced with /nt/, which occur less frequently. We conducted two experiments in order to investigate how exposure and phonetic context influence this reported processing disadvantage for flapped variants. Experiment 1 demonstrated that the time to recognize flapped variants presented in isolation decreased over the course of the experiment, while accuracy increased. Experiment 2 replicated this finding and showed further that flapped variants that were presented in a casually produced sentence context were recognized faster compared with flapped variants presented in a carefully produced sentence context. Interestingly, the effect of context emerged only in late responses and was present only for flapped but not for canonical variants. Our results thus show that increased exposure and phonetic context help listeners recognize allophonic variants. This finding provides further support for the notion that listeners are flexible and adapt to phonetic variation in speech.

Keywords Speech reduction · Spoken word recognition · Allophonic variation · Flapping

Spoken language is characterized by an extraordinary amount of variation. For example, the way words are produced in conversational speech often differs from the ideal or *canonical* pronunciations that we find in dictionaries, due, in part, to acoustic-phonetic reductions that lead to the shortening of speech segments or the deletion of phonemes and even whole syllables (e.g., Johnson, 2004). For example, the vowel schwa is often shortened or deleted, as when the word *support* is pronounced as *s'pport* (Patterson, LoCasto, & Connine, 2003). Another segment that is often reduced is word-final /t/, which is frequently deleted or produced as a glottal stop, as when the word *mist* is pronounced as *mis* (see, e.g., Ernestus & Warner, 2011, for more examples). Another source of variability in the pronunciation of words is allophonic variation. Allophonic variation refers to articulatory and acoustic differences among speech sounds

belonging to the same phonemic category (Ladefoged, 2000). In American English, one such allophonic process is nasal flapping. A flap (or tap) is a rapid movement of the tip of the tongue upward to contact the roof of the mouth (Ladefoged, 2000). In nasal flapping, sound sequences consisting of an alveolar nasal and an alveolar oral stop (/nt/ and /nd/) become nasalized voiced flaps (phonetically transcribed as /ɾ/) when they occur between a stressed and an unstressed vowel. For example, the word *center* can either be pronounced as /sɛntə/ or as /sɛɾə/. Listeners thus encounter two different phonetic variants of one word: a canonical variant that is produced with /nt/ and a flapped variant that is produced with a nasal flap. So far, relatively little is known about how listeners process allophonic variation. Acoustic-phonetic variability in general and allophonic variation in particular pose a challenge to psycholinguistic theories that attempt to describe the processes and representations that allow language users to produce and comprehend spoken language (e.g., Luce & McLennan, 2005).

The goal of the present study was to investigate how listeners recognize flapped variants and how the ability of listeners to adapt to variation in speech helps in this task. In particular, we examined two ways in which listeners may adapt to this kind of allophonic variation: First, we

✉ Malte C. Viebahn
malte.viebahn@uni-leipzig.de

¹ Department of Psychology, University of Leipzig, Neumarkt 9-19, 04109 Leipzig, Germany

² Department of Psychology, University at Buffalo, Buffalo, NY, USA

investigated whether increased exposure to flapped variants influences how listeners recognize flapped words. Second, we examined whether listeners use information about the phonetic context in which the flapped variants are embedded in order to recognize words with flaps.

Analyses of speech corpora have shown that flapped production variants are more common in American English than canonical production variants (e.g., Ranbom & Connine, 2007). However, studies that investigated perceptual processing of flapped variants have found that words that undergo nasal flapping are recognized more slowly and less accurately compared with canonical variants (Pitt, 2009; Ranbom & Connine, 2007). These studies have also shown that the speed with which flapped variants are recognized depends on how frequently they occur. Highly frequent variants are recognized more easily and more accurately than less frequent variants. However, even when controlling for variant frequency, flapped variants appear to be recognized less rapidly than canonical variants. This finding is surprising because it suggests a contradiction to the classic word frequency effect, according to which word forms that occur more frequently are associated with faster and more accurate recognition performance (for a review, see Luce & Pisoni, 1998). Nasal flapping is thus a particularly interesting kind of allophonic variation to examine because the frequency of occurrence of the variant that is less well recognized (i.e., the flapped variant) is much higher compared with the more easily recognized but less frequent canonical variant.

Various hypotheses have been put forth about the source of this processing disadvantage for flapped variants. For example, Pitt (2009) argues that canonical forms may be easier to recognize because they are perceptually more distinct given that they contain more phonetic information. The more phonetic information a word form contains, the less likely it will be confused with a different word and the more easily it can be recognized. A similar proposal is that flapped variants are more difficult to recognize because they reside in larger lexical neighborhoods. Spoken words with large lexical neighborhoods are recognized more slowly because they compete with other similar sounding words (Luce & Pisoni, 1998). Because shorter words tend to have larger lexical neighborhoods than longer words and flapped variants are shorter than canonical variants, it is plausible to assume that flapped variants have more lexical competitors than canonical variants. As a consequence, lexical inhibition due to denser neighborhoods may neutralize any advantage due to higher production frequencies and thus decrease recognition performance. Another possibility is that the canonical processing advantage stems from increased overlap with orthographic forms (Ranbom & Connine, 2007). Several studies have suggested that the amount of overlap between phonological and orthographic forms can influence how spoken words are recognized (e.g., Grainger, Muneaux, Farioli, & Ziegler, 2005; Perre,

Pattamadilok, Montant, & Ziegler, 2009; Seidenberg & Tanenhaus, 1979; Taft, Castles, Davis, Lazendic, & Nguyen-Hoan, 2008; but also see Cutler, Treiman, & van Ooijen, 1998; Mitterer & Reinisch, 2015). Canonical forms may therefore be recognized more easily compared with flapped forms because the /nt/ sequence in the canonical variant is explicitly represented in the orthography whereas the nasal flap in the flapped variant is not.

Another proposal is that listeners typically engage in different modes of processing when listening to casual speech (which consists primarily of reduced pronunciation variants) compared with when listening to careful speech (in which canonical variants are more likely to occur). For example, according to Sumner (2013), the processing disadvantage for flapped variants reported in previous experiments stems from the fact that the recognition of reduced speech involves more top-down processing than the recognition of canonical speech, which is presumably based more on bottom-up processing. This proposal is supported by evidence showing that canonical and flapped forms yielded equivalent priming effects in semantic and long-term priming experiments (Sumner, 2013). Further evidence for the notion that the recognition of reduced word forms depends on the processing mode of the listener comes from studies showing that task difficulty can influence the way in which allophonic variants are recognized (McLennan, Luce, & Charles-Luce, 2003). In a series of long-term priming experiments, McLennan et al. (2003) found that variant-specific priming effects emerge only when the task is relatively easy and allows for shallow processing. These studies suggest that the way listeners recognize allophonic variants depends on how quickly they process linguistic input.

A related proposal claims that differences in processing performance are due to differences in how flapped and canonical forms are encoded during word-form acquisition (Sumner, Kim, King, & McGowan, 2014). According to this theory, canonical forms are acquired in socially salient situations in which the language learner's attention is actively directed toward the acoustic-phonetic detail of how words are supposed to be pronounced. This type of learning situation yields very fine-grained phonetic representations. In contrast, reduced word forms acquired in situations in which the precise phonetic form is less salient, which results in coarser, less detailed representations. As a consequence, the more detailed albeit less frequently activated representations of the canonical variants are recognized better compared with the more frequently activated but phonetically less detailed representations of the reduced variants.

The present study examines another factor that might influence the recognition of flapped variants: listeners' ability to adapt to variation in speech. In previous studies, listeners may have processed flapped variants more slowly than canonical variants because they did not have enough time or enough

information available to them in order to adapt to variation in speech. Research has shown that the speech recognition system is quite flexible and that listeners are capable of adapting to acoustic-phonetic variation from a variety of sources. For example, Clarke and Garret (2004) showed that American English listeners can adapt to Spanish-accented and Chinese-accented speech after only a minute of exposure. Similarly, Dutch listeners have been shown to adapt to Hebrew-accented Dutch even if they had had only limited experience with this accent (e.g., Witteman, Weber, & McQueen, 2013). Evidence for listeners' ability to adapt to variable speech input comes also from research on perceptual learning in speech (e.g., Clarke-Davidson, Luce, & Sawusch, 2008; Connine & Darnieder, 2009; Kraljic, Brennan, & Samuel, 2008; Norris, McQueen, & Cutler, 2003). These studies have shown that listeners adapt their phonemic perception after being exposed to ambiguous speech sounds in lexically biasing contexts. Clearly, if listeners adapt to ambiguous speech sounds and to variation arising from foreign-accented speech, they also ought to be able to adapt to different pronunciation variations within their own native language. In the present study, we investigated listeners' ability to adapt to nasal flapping by measuring the processing advantage for canonical variants over the course of a lexical-decision experiment and by presenting phonetic context that was produced either in a matching or mismatching speaking style.

Previous studies have shown that contextual information can help listeners recognize variable speech input. For example, several studies have shown that preceding semantic and syntactic information can facilitate the recognition of reduced pronunciation variants (e.g., Ranbom & Connine, 2007; Tuinman, Mitterer, & Cutler, 2014; van de Ven, Tucker, & Ernestus, 2011; Viebahn, Ernestus, & McQueen, 2015). Another way in which listeners may use context to adapt to flapped variants is by using phonetic information. Previous work has shown that phonetic information can influence the recognition of massively reduced word forms. For example, Ernestus, Baayen, and Schreuder (2002) demonstrated that Dutch listeners' ability to correctly identify strongly reduced word forms that had been spliced out of connected speech improved dramatically when context was provided. When the reduced forms were presented in isolation (e.g., [mok], a reduced form of Dutch /mox l k/ "possible"), recognition accuracy was only 52%. In contrast, when the forms were presented together with neighboring segments (e.g., [ɛlmokna]), accuracy increased to 70%. Local phonetic context has also been shown to be important in the recognition of segmental reductions which occur in specific phonological environments such as nasal place assimilation and word-final /t/ deletion (e.g., Coenen, Zwitserlood, & Boelte, 2001; Gaskell & Marslen-Wilson, 1998; Mitterer & Ernestus, 2006; Mitterer & McQueen, 2009).

In addition to local phonological environments, phonetic context may also have more global effects on speech recognition. For example, exposure to casual speech with multiple reduced word forms decreases phonological competition effects in visual-world eye-tracking experiments (Brouwer, Mitterer, & Huettig, 2012). Furthermore, in an EEG study using Turkish-accented Dutch, Hanulíková, van Alphen, van Goch, and Weber (2012) showed that the way morpho-syntactic violations are processed by native listeners is influenced by their knowledge of the accent of the speaker. Viebahn, Ernestus, and McQueen (2017) obtained a similar result for morpho-syntactic violations that can occur in casually produced Dutch. In particular, the authors demonstrated that schwas that serve as grammatical markers were perceived as grammatical errors when the speaker talked with a careful speaking style. But when the speaker spoke in a casual style, listeners appeared to attribute the absent schwa to the speaking style rather than interpreting it as a grammatical error. These studies show that the way in which listeners process acoustic-phonetic variation depends on the phonetic context, suggesting that listeners use global phonetic information in order to adapt to reduced speech.

A related question is *when* context influences recognition. More specifically, whether context is used to *predict* word-form variants or whether context influences recognition at later stages of processing. There is a growing body of research showing that listeners can use several sources of information in order to predict upcoming linguistic input (Altmann & Kamide, 1999; Arai & Keller, 2013; Kamide, Scheepers, & Altmann, 2003; van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). In particular, a number of studies demonstrate that listeners can predict phonological information about words based on semantic information (e.g., Dahan & Tanenhaus, 2004; DeLong & Kutas, 2008; Ito, Pickering, & Corley, 2018), syntactic information (Magnuson, Tanenhaus, & Aslin, 2008), and regional accent (Brunelliere & Soto-Faraco, 2013). There is currently a debate concerning whether prediction is a general cognitive mechanism that is not restricted to language processing (Clark, 2013; Friston, 2010). However, studies suggest that whereas listeners are often able to make predictions about language input, prediction may be less common than suggested and dependent on the listening situation and characteristics of the listener (e.g., Fine, Jaeger, Farmer, & Qian, 2013; Huettig & Mani, 2016; Viebahn et al., 2015). In the present study, we address the issue of context-based prediction by examining the time course of phonetic context effects in the recognition of flapped pronunciation variants.

The current study will focus on the recognition of allophonic variants that are produced with a nasal flap instead of the canonical /nt/ sequence. We explore two ways in which listeners may adapt to this form of variation: Experiment 1 investigates if increased exposure to nasal flapping allows

listeners to adapt to flapped word forms. If this is the case, the previously reported processing advantage for canonical over flapped pronunciation variants ought to decrease over the course of the experiment. Experiment 2 examines whether listeners can use global phonetic context in order to adapt to flapped pronunciation variants. To this end, canonical and flapped pronunciation variants were presented in casually or carefully produced carrier sentences that are semantically neutral. If listeners are sensitive to global phonetic context information, flapped variants ought to be processed more easily when the preceding context is produced in a casual manner than when it is produced in a careful manner. We examine the time course of this context effect by analyzing distributions of reaction times (RTs): If context is used in order to predict the allophonic variant, we predicted a simple shift in the RT distributions. In contrast, if context is used at a later stage, we predicted that the effect of context will influence primarily slow responses in the tail of the distributions.

Another question about the influence of phonetic context is whether contextual information influences both variants equally or whether context is only used for the recognition of flapped variants. According to episodic theories of lexical representation (e.g., Goldinger, 1998), flapped and canonical variants are stored in the mental lexicon, and neither of them has a privileged status. The degree to which they become activated is primarily a function of their frequency of occurrence and the degree to which they overlap with the phonetic input. In contrast, according to abstractionist and hybrid accounts, the canonical form has a privileged status. According to the abstractionist view (e.g., Gaskell & Marslen-Wilson, 1998; Norris & McQueen, 2008), the canonical form is the only form that is represented in memory, and all other forms are recognized via some kind of sublexical process (e.g., phonological inference or perceptual learning). In contrast, hybrid accounts (e.g., McLennan et al., 2003; Ranbom & Connine, 2007) assume that there can be more than one lexical representation. However, the representation of the canonical form has a privileged status either because of the way in which this form was encoded during initial acquisition (Sumner et al., 2014) or because of its overlap with its orthographic form (Ranbom & Connine, 2007). These approaches make different prediction with respect to the influence of contextual information. Whereas the episodic account should predict that both the canonical and the flapped variants will benefit from matching phonetic context, abstractionist and hybrid accounts predict that context effects will be stronger for flapped compared with canonical variants.

Experiment 1: Adaptation to flapped variants in isolation

The first experiment investigated if listeners adapt to flapped pronunciation variants over time. As was done

in previous studies that demonstrated a processing advantage for canonical over flapped pronunciation variants (e.g., Ranbom & Connine, 2007), we used the lexical decision task in order to measure recognition performance. We predicted that the results will show the effect of variant type previously found: Overall, flapped variants ought to be recognized more slowly and less accurately compared with canonical variants. Furthermore, because previous studies have shown that listeners can adapt to variation in speech, we predicted that the processing advantage for canonical variants would decrease or even disappear over the course of the experiment, demonstrating that listeners can adapt to this type of allophonic variation.

Method

Participants

Participants were 60 students of the University at Buffalo. All were native speakers of American English, and none reported a history of speech or hearing problems.

Materials

We used the 48 words with an intersyllabic /nt/ sequence and the 48 filler words without an intersyllabic /nt/ sequence reported in Ranbom and Connine (2007). Frequency estimates based on Kučera and Francis (1967) ranged from zero to 331 per million for the /nt/ words (mean: 28.23, standard deviation: 59.97) and from zero to 116 for the filler words (mean: 17.9, standard deviation: 23.63). In addition, 48 nonwords with and 48 nonwords without an intersyllabic /nt/ sequence were created. An additional six words without /nt/ sequence served as practice items (frequency range: five to 377, mean: 77.33, standard deviation: 147.5) along with six more nonwords. These materials are listed in the Appendix. The words and nonwords were embedded at the end of the carefully produced carrier phrase *I want to say . . .* and the casually produced carrier phrase *I wanna say . . .* and recorded in a sound-attenuated room by a male native speaker of a Midwestern dialect of American English. After the recordings were made, the words were excised from the carrier sentence so that they could be presented in isolation. As an example, annotated waveforms and spectrograms for a canonical and a flapped variant of the word *center* are shown in Fig. 1.

The average duration of the canonical variants was 609 ms ($SD = 113$ ms), whereas the average duration of the flapped variants was 441 ms ($SD = 79$ ms). In order to ensure that differences in RTs in the lexical-decision experiment would not be influenced by differences in stimulus duration, we

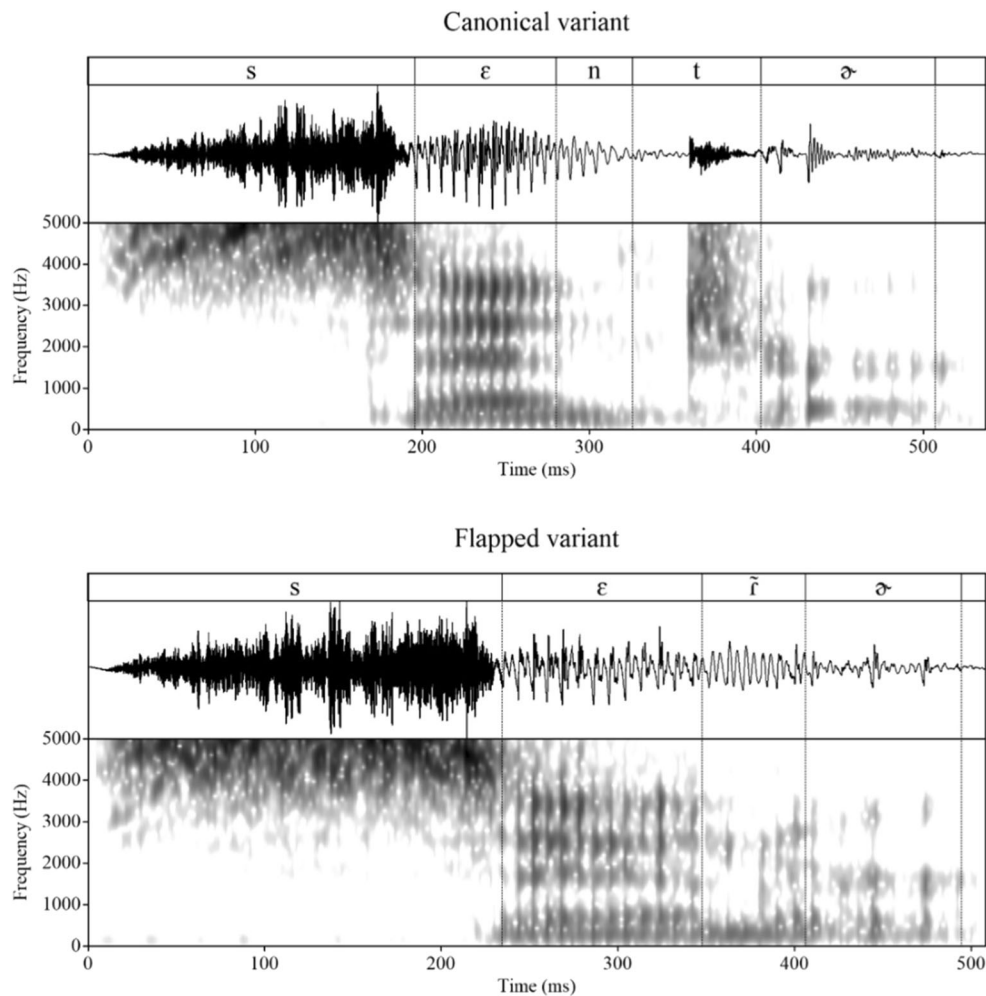


Fig. 1 Waveforms and spectrograms of the canonical and the flapped pronunciation variants of the word *center*, which was used as a stimulus in Experiment 1

equated the durations of the two variants (canonical and flapped) for each word type using Praat's overlap-add algorithm (Boersma & Weenink, 2010), which is an implementation of the TD-PSOLA method (Moullies & Charpentier, 1990). This was done in the following way: First, we measured the difference in duration between the two variants, which was on average 168 ms ($SD = 66$ ms). Second, the longer token (the canonical variant) was shortened by half this amount (on average: 84 ms), whereas the shorter token was lengthened by the same amount. The same procedure was used for the flapped and canonical variants of the nonwords which contained an /nt/ sequence. For the words and nonwords that did not contain an /nt/ sequence, lengthened and shortened versions were created by changing the durations by constant ratios for each stimulus group. This procedure ensured that all tokens were changed and that the change would only be relatively small. Each sound file was checked manually to ensure that the change was only barely noticeable. The original and adjusted sound files of the flapped and canonical words can be downloaded here (<https://osf.io/3tmbr/>).

Procedure

For the presentation of the stimuli, two lists were created. In each list, half of the words and nonwords with an intersyllabic /nt/ sequence were produced with a nasal flap, whereas the other half was produced with a canonical /nt/ sequence. An equal number of participants was run for each list. Participants were tested individually or in groups of up to three people in a sound-attenuated room. They were seated at a desk with an Apple Macintosh computer, a button box, and a pair of headphones. Presentation of the stimuli, randomization, and response collection was controlled by the PsyScope software package (Cohen, MacWhinney, Flatt, & Provost, 1993). Participants were instructed to listen to the stimuli presented to them over headphones and then to respond *word* with their right hand or *not a word* with their left hand as quickly and as accurately as possible by pressing appropriately labeled buttons. Each trial proceeded as follows: A black fixation cross was presented before a white background in the center of the screen for 500 milliseconds. A word or nonword was then

presented over headphones, and the participant had 5 seconds to provide a button-press response. If the participant did not make a response after the 5 seconds had expired, the computer automatically presented the next trial. The intertrial interval was 500 milliseconds. Before the presentation of the experimental stimuli, 12 practice trials were presented in order to familiarize the participants with the task. No feedback regarding the accuracy of the participants' responses was given.

Results

In order to investigate whether recognition performance changed over the course of the experiment, trials were divided into two equal-sized parts. The first part ranged from Trial 1 to Trial 96, and the second part ranged from Trial 97 to Trial 192. Note that for the calculation of these trial numbers, the number of practice trials was subtracted. RT was measured from the acoustic onset of the word. Mean RTs and accuracy rates for each part and variant are shown in Table 1, split for words and nonwords.

We used generalized linear mixed models with a binomial link function for the analysis of the accuracy rates and linear mixed-effects models for the analysis of the RTs (Baayen, Davidson, & Bates, 2008). Analyses were carried out using the R programming language (R Development Core Team, 2007) and the lme4 package for linear mixed-effects models (Bates, 2005). The models were fitted using the *bobyqa* optimizer, and the maximum number of iterations was set to 20,000. For linear mixed-effects models, *p* values were estimated by using the standard normal distribution. The main predictor variables in the model were variant type (canonical vs. flapped), trial, and the interaction between variant type and trial. For the RT models, the RT on the previous trial was added as a control variable to improve model fit (Baayen & Milin, 2010). The models included random intercepts for participants and items, and random slopes for variant type, trial, the interaction term, and RT on the previous trial. In other words, models were fitted using the maximum random effects structure (Barr, Levy, Scheepers, & Tily, 2013). Correlations

between the random effects were not estimated in order to facilitate the fitting procedure. The RT variables and the trial variable were *z* standardized for the purposes of the modeling procedure, and the variable variant type was contrast coded (canonical variant = -0.5 , flapped variant = 0.5).

For the analysis of the RTs, we followed the recommendations of Baayen and Milin (2010), who propose a combination of two types of outlier cuts. First, we excluded extreme values that are likely due to participants not having performed the task correctly (either because they pressed a button by accident which may result in extremely fast RTs, or because they responded so slowly that it is unlikely that their responses reflect online processing). This cut was applied to both the accuracy and the RT data and was performed before any statistical analyses were conducted. We regarded trials with response times less than 200 ms or greater than 2,500 ms as extreme values and removed them from the analysis (1.2% of responses). Second, we excluded overly influential outliers during the statistical modeling of the RT data. This refers to data points with absolute standardized residuals exceeding 2.5 standard deviations. The percentage of removed outliers varied across models and ranged from 2.61% to 3.55%.

Accuracy

As shown in Table 1 and Fig. 2, task performance was quite good overall, but accuracy rates were clearly lower for flapped compared with canonical pronunciation variants. Accuracy for flapped variants improved by approximately 5 percentage points over the course of the experiment. As summarized in Table 2, the statistical analysis showed a significant main effect of variant type and a significant interaction between variant type and trial. Separate models for each pronunciation variant showed a significant increase in accuracy over the course of the experiment for flapped variants ($\beta_{\text{Trial}} = 0.24$, $SE = 0.10$, $z = 2.49$, $p = .01$) but not for canonical variants ($\beta_{\text{Trial}} = -0.11$, $SE = 0.13$, $z = -0.87$, $p = .38$). Separate analyses for each of the two parts of the experiment indicated that the difference in accuracy between the two variants remained significant over the course of the experiment.

Table 1 Mean RTs and accuracy rates for Experiment 1

	Item type	RT				Accuracy			
		Mean	Part 1	Part 2	Difference	Mean	Part 1	Part 2	Difference
Words	Canonical	924 (12)	929(17)	920 (16)	−9	91.17 (0.90)	91.29 (1.28)	91.05 (1.27)	−0.24
	Flapped	997 (14)	1,030 (22)	964 (17)	−66	75.32 (1.43)	72.89 (2.11)	77.75 (1.90)	4.86
	Other	967 (11)	967(16)	966 (16)	−1	89.75 (0.72)	91.45 (0.91)	88.06 (1.09)	−3.39
Nonwords	Canonical	1,012 (16)	1,033 (24)	990 (22)	−43	91.47 (1.20)	91.47 (1.58)	91.47 (1.82)	0.00
	Flapped	1,032 (16)	1,058 (24)	1,005 (21)	−53	87.87 (1.21)	88.71 (1.65)	87.02 (1.77)	−1.69
	Other	1,070 (16)	1,082 (26)	1,059 (18)	−23	92.51 (1.12)	92.50 (1.69)	92.52 (1.47)	0.02

Note. RTs are given in milliseconds and accuracy rates as percentages. Standard errors are shown in parentheses

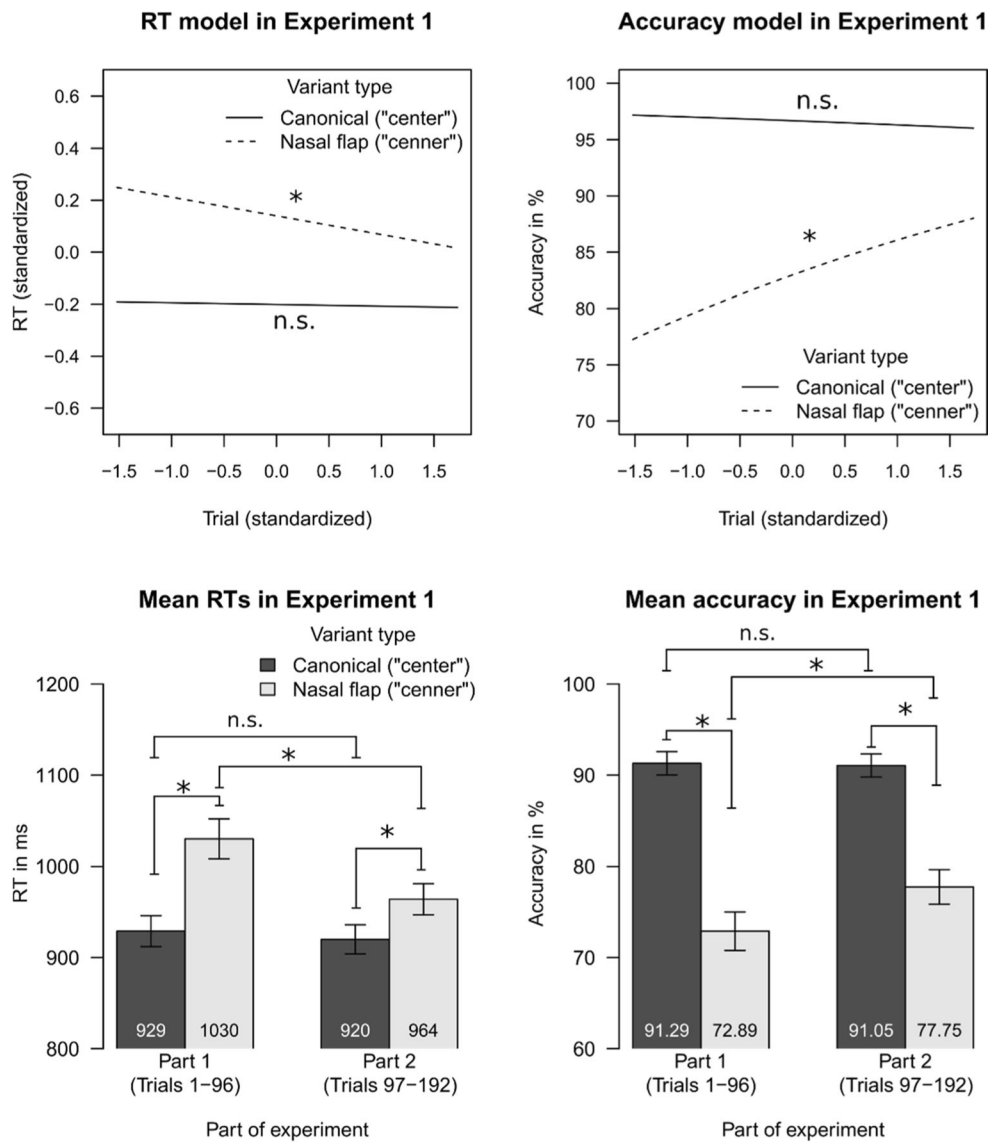


Fig. 2 Results of Experiment 1. * = statistically significant effect; n.s. = absence of a significant effect. Error bars indicate one standard error around the mean

Table 2 Model parameters for mixed-effects models for RTs and accuracy rates for Experiment 1

	Accuracy				RT			
	β	SE	z value	p value	β	SE	t value	p value
<i>Fixed effects</i>								
Intercept	2.48	0.24	10.50		-0.03	0.07	-0.45	
Variant type	-1.79	0.26	-6.85	<.001	0.34	0.05	7.61	<.001
Trial	0.06	0.09	0.74	.46	-0.04	0.22	-1.77	.08
Previous RT	-	-	-	-	0.16	0.03	6.39	<.001
Variant Type \times Trial	0.35	0.16	2.21	.03	-0.07	0.03	-2.19	.03
<i>Random effects</i>								
	$SD_{\text{participants}}$	SD_{words}			$SD_{\text{participants}}$	SD_{words}		
Intercept	0.72	1.30			0.34	0.34		
Variant type	0.49	1.29			0.11	0.23		
Trial	0.26	0.19			0.12	0.06		
Previous RT	-	-			0.12	0.07		
Variant Type \times Trial	0.27	0.26			0.00	0.08		

Note. Trial numbers and RTs were standardized before being entered into the model. SE = standard error; SD = standard deviation

However, as indicated by the beta values, the size of the difference was larger in Part 1 ($\beta_{\text{Variant}} = -2.0$, $SE = 0.32$, $z = -6.30$, $p < .001$) compared with Part 2 ($\beta_{\text{Variant}} = -1.49$, $SE = 0.31$, $z = -4.78$, $p < .001$). In order to explore whether adaptation to flapped variants is influenced by the frequency with which these variants occur, we ran an additional model for the 28 words for which Ranbom and Connine (2007) reported frequency counts from the Switchboard corpus. This model showed no interaction between variant type, trial, and log-transformed nasal-flap frequency ($\beta_0 = 2.44$, $SE = 0.25$, $z = 9.8$; $\beta_{\text{Variant} \times \text{Trial} \times \text{Variant Frequency}} = 0.12$, $SE = 0.27$, $z = 0.43$, $p = .67$).

In order to examine whether recognition performance for the filler words and nonwords improved, we ran additional models for each stimulus type. There was a significant decrease in accuracy for the filler words ($\beta_{\text{Trial}} = -0.25$, $SE = 0.07$, $z = -3.41$, $p < .001$). In contrast, the model showed no significant influence of trial number on accuracy for the nonwords with a canonical /nt/ sequence ($\beta_{\text{Trial}} = 0.06$, $SE = 0.18$, $z = 0.31$, $p = .76$), with a nasal flap ($\beta_{\text{Trial}} = 0.03$, $SE = 0.12$, $z = 0.23$, $p = .82$), or with no /nt/ sequence ($\beta_{\text{Trial}} = 0.05$, $SE = 0.10$, $z = 0.51$, $p = .61$).

Reaction times

As shown in Table 1 and Fig. 2, the overall RTs for the flapped variants were slower compared with the canonical variants. However, the RT difference between the two variants decreased over the course of the experiment. As shown in Table 2, the analyses showed a significant interaction between variant type and trial. Separate models for each pronunciation variant showed a significant decrease of RTs over the course of the experiment for flapped variants ($\beta_{\text{Trial}} = -0.07$, $SE = 0.03$, $t = -2.48$, $p = .01$) but not for canonical variants ($\beta_{\text{Trial}} = -0.01$, $SE = 0.02$, $t = -0.59$, $p = .56$). In order to examine the time course of the variant-type effect, separate models were fitted for each of the two parts of the experiment. These models showed that listeners responded more slowly to flapped compared with canonical variants in the first part ($\beta_{\text{Variant}} = 0.33$, $SE = 0.05$, $t = 7.02$, $p < .001$) and the second part of the experiment ($\beta_{\text{Variant}} = 0.28$, $SE = 0.06$, $t = 4.76$, $p < .001$). However, as indicated by the beta values, the processing advantage for canonical over flapped variants decreased over the course of the experiment. A model based on the 28 words for which nasal-flap frequency counts were available showed no interaction between variant type, trial, and flap frequency ($\beta_0 = -0.07$, $SE = 0.09$, $t = -0.73$; $\beta_{\text{Variant} \times \text{Trial} \times \text{Variant Frequency}} = -0.05$, $SE = 0.04$, $t = -1.26$, $p = 0.21$), suggesting that adaptation was not influenced by variant frequency.

Additional models for the filler words and nonwords showed no significant influence of trial number on the filler words ($\beta_{\text{Trial}} < .001$, $SE = 0.02$, $t = 0.02$, $p = .98$). In contrast,

RTs to nonwords with an /nt/ sequence ($\beta_{\text{Trial}} = -0.07$, $SE = 0.03$, $t = -2.77$, $p = .01$), with a nasal flap ($\beta_{\text{Trial}} = -0.09$, $SE = 0.02$, $t = -3.80$, $p < .001$), and with no /nt/ sequence ($\beta_{\text{Trial}} = -0.07$, $SE = 0.03$, $t = -2.63$, $p = .01$) decreased significantly over the course of the experiment.

Discussion

The results of Experiment 1 demonstrate that listeners adapted to flapped pronunciation variants over the course of a lexical-decision experiment. Listeners' response times and accuracy rates for flapped variants improved during the experiment, whereas response times and accuracy to canonical variants remained constant. These results show that the processing disadvantage decreased considerably toward the end of the experiment, suggesting that the processing advantage for canonical over flapped variants is smaller than previously assumed (e.g., Pitt, 2009; Ranbom & Connine, 2007). Similar to the canonical variants, the filler words also showed no improvement during the experiment. In contrast, the analyses of the nonword responses showed an increase in speed, suggesting that listeners became more familiar with nonwords. Crucially, however, the finding that only word responses to flapped variants improved over the course of the experiment indicates that listeners adapted specifically to these kinds of words rather than adapting to the task in general. While the accuracy and RT results show an improvement in the correct recognition of the flapped variants, we observed a significant difference between responses to canonical and flapped variants even at the end of the experiment, suggesting that the canonical forms remained somewhat easier to recognize than flapped variants. (Of course, this difference may have been further reduced and perhaps eliminated with more exposure to the variants.) Our results thus demonstrate that listeners are flexible with regard to how they recognize pronunciation variants, which is consistent with findings showing that listeners are flexible when listening to foreign-accented speech. We discuss potential theoretical explanations regarding the mechanisms that allow listeners to adapt to allophonic variation in the General Discussion. First, however, we will investigate whether listeners can also use global phonetic context in order to adapt to nasal flapping.

Experiment 2: Words in phonetic context

The goal of Experiment 2 was to examine whether listeners would benefit from phonetic context information such as a casual speaking style when recognizing flapped pronunciation variants. We used the same tokens as in Experiment 1, but this time presented them in a carrier sentence that

was either produced in a careful or a casual manner. As a carrier sentence, we chose *I want to say . . .*, which also contains an intersyllabic /nt/ sequence that can be produced as a nasal flap (the words *want to* can be produced as *wanna*). The careful version of the carrier sentence was *I want to say . . .*, whereas the casual version was *I wanna say . . .*. Previous studies have shown that the processing of word-form variation due to acoustic-phonetic reduction or accented speech is improved by phonetic context (e.g., Brunelliere & Soto-Faraco, 2013; Ernestus et al., 2002). Since nasal flapping is a type of word-form variation, we predicted that the recognition of flapped variants will improve when these variants are presented in a matching phonetic context (i.e., when they are presented in a casually compared with a carefully produced carrier phrase). Moreover, we predicted that if the canonical form has a privileged status in the mental lexicon, flapped variants ought to benefit more from matching phonetic context than canonical forms. In contrast, if neither of the variant types has a privileged status, then the effect of context ought to be the same for both variants. In order to examine the time course of the effect of context, we analyzed the RT distributions. If context is used in order to predict the allophonic variant, we predicted a simple shift in the RT distributions. In contrast, if context is used at a later stage, we predicted that the effect of context would influence primarily slow responses rather than fast responses. Furthermore, we examined whether listeners' recognition of flapped variants improved over the course of the experiment in order to replicate the main result of Experiment 1.

Method

Participants

Participants were 48 students of the University at Buffalo who had not participated in Experiment 1. All were native speakers of American English, and none reported a history of speech or hearing problems.

Materials

The materials were based on the same recordings as in Experiment 2. The words and nonwords were combined with a recording of the carrier phrase *I want to say . . .*. This carrier phrase was either produced in a careful manner (*I want to say . . .*) or in a casual manner (*I wanna say . . .*), such that there were four versions for each item: a canonical variant in a careful carrier phrase, a canonical variant in a casual carrier phrase, a flapped variant in a careful carrier phrase, and a flapped variant in a casual carrier phrase. For each word and nonword without an /nt/ sequence, two versions were created: one in the casual and one in the careful carrier phrase. The

duration of the careful carrier sentence was 1,459 ms, and the duration of the casual carrier sentence was 789 ms. Figure 3 shows a stimulus example for each of the four conditions.

Procedure

The versions of each item were distributed across four lists so that participants would hear each word and nonword only once. Each list consisted of 48 experimental words, 48 nonwords with an /nt/ sequence, 48 filler words, and 48 nonwords without an /nt/ sequence. Half of the words and nonwords occurred in the casual and the other half in the careful carrier phrase. Of the words and nonwords that contained an /nt/ sequence, half were produced canonically and half were produced with a nasal flap. The procedure of the lexical-decision task was identical to the task in Experiment 1, with the exception that participants were told that they would hear short sentences rather than isolated words and that they should respond to the last word or nonword in the sentence.

Results

We used the same statistical methods as in Experiment 1. As in Experiment 1, trials with response times less than 200 ms or greater than 2,500 ms were regarded as extreme values and removed from the analysis (1% of responses). Furthermore, for the RT analyses data points with absolute standardized residuals exceeding 2.5 standard deviations were removed from the statistical model. Based on this criterion, the number of removed data points ranged from 1.33% to 4.38%. Continuous variables were z standardized. The binary predictor variables were contrast coded: The canonical variant was coded as -0.5 , whereas the flapped variant was coded as 0.5 . Similarly, the careful context was coded as -0.5 and the casual context as 0.5 . In order to examine the time course of the effect of context, we examined whether the effect of context differed depending on overall response speed. We divided the data into “fast” and “slow” responses. Fast responses were defined as the bottom 40%, and slow responses as the upper 60%. The exact cut-off point was based on the visual inspection of the RT distributions. For the statistical analysis, responses categorized as slow were contrast coded as -0.5 , and responses categorized as fast were coded as 0.5 .

Accuracy

Table 3 shows RTs and accuracy scores split by variant type, part of experiment, and word/nonword status. Overall accuracy was similar to Experiment 1. There was an increase in accuracy for flapped variants from the first to the second part of the experiment of approximately 8 percentage points. In contrast, accuracy changes for the other stimulus types were negligible. When looking at the influence of phonetic context (see Table 4), we see a similar pattern. Accuracy for flapped

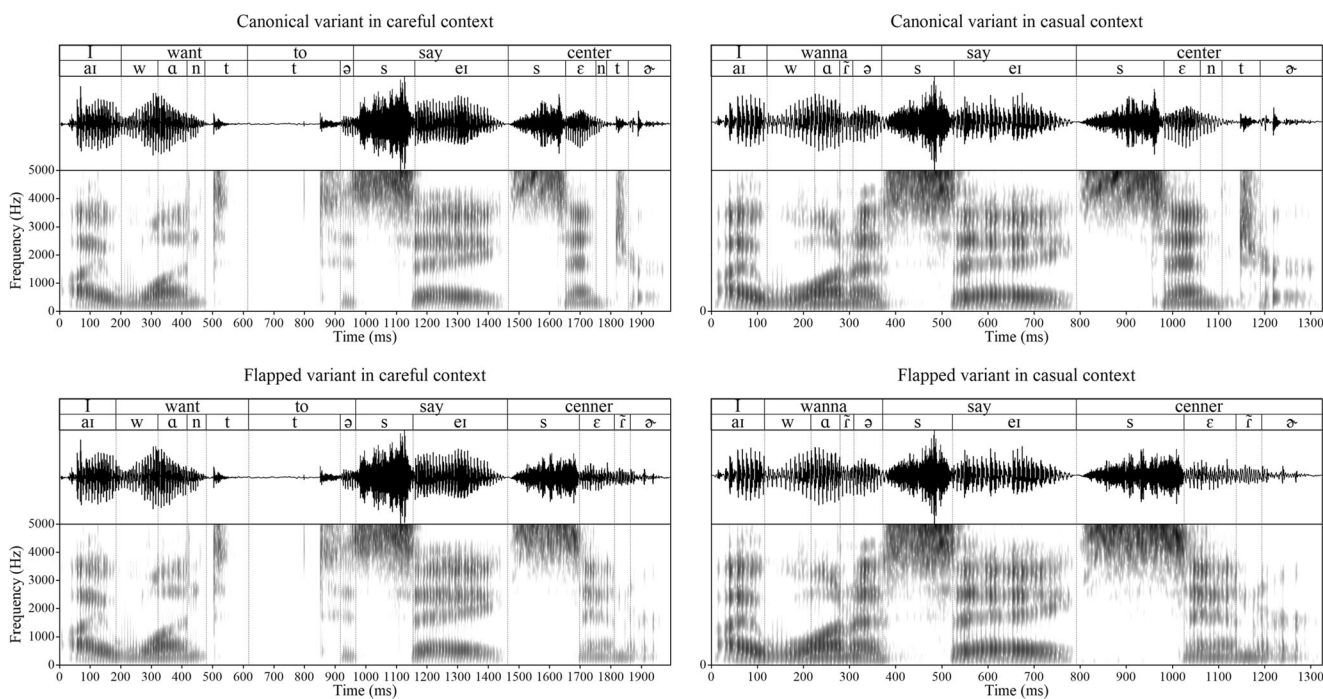


Fig. 3 Stimulus examples for the four conditions in Experiment 2

variants is slightly higher (approximately 4 percentage points) in the casual compared with the careful context. In contrast, accuracy for the other stimulus types appears to be unaffected by phonetic context.

Table 5 shows the results of a model including a three-way interaction between variant type, context, and response speed as well as a two-way interaction between variant type and trial. According to this model, there was no influence of context, but a significant interaction between variant type and trial. Separate models for each pronunciation variant showed a significant increase in accuracy over the course of the experiment for flapped variants ($\beta_{\text{Trial}} = 0.05, SE = 0.10, z = 4.85, p < .001$) and a decrease for canonical variants ($\beta_{\text{Trial}} = -0.25, SE = 0.13, z = -1.95, p = .05$). Separate analyses for each of the two parts of the experiment

indicated that the difference in accuracy between the two variants remained significant over the course of the experiment. However, as indicated by the beta values, the size of the difference was larger in Part 1 ($\beta_{\text{variant}} = -2.28, SE = 0.35, z = -6.49, p < .001$) compared with Part 2 ($\beta_{\text{variant}} = -1.28, SE = 0.34, z = -3.76, p < .001$).

In order to examine whether recognition performance for the filler words and nonwords improved, we ran additional models for each stimulus type. These models showed no significant influence of trial number on accuracy for the filler words ($\beta_{\text{Trial}} = -0.11, SE = 0.10, z = -1.05, p = .29$), the nonwords with a canonical /nt/ sequence ($\beta_{\text{Trial}} = 0.17, SE = 0.18, z = 0.92, p = .36$), with a nasal flap ($\beta_{\text{Trial}} = 0.19, SE = 0.13, z = 1.42, p = .16$), or with no /nt/ sequence ($\beta_{\text{Trial}} = 0.21, SE = 0.11, z = 1.90, p = .06$).

Table 3 Mean RTs and accuracy rates for Experiment 2, split by variant type, part of experiment, and word/nonword status

	Item type	RT				Accuracy			
		Mean	Part 1	Part 2	Difference	Mean	Part 1	Part 2	Difference
Words	Canonical	917 (14)	930(22)	904 (19)	-26	90.75 (0.81)	92.35 (0.99)	89.15 (1.24)	-3.20
	Flapped	995 (18)	1,017 (28)	973 (24)	-44	75.01 (1.59)	70.93 (2.34)	79.08 (2.00)	8.15
	Other	959 (15)	956(21)	962 (22)	6	93.02 (0.64)	93.27 (0.86)	92.77 (0.96)	-0.50
Nonwords	Canonical	1,037 (23)	1,069 (39)	1,005 (25)	-64	90.88 (1.91)	90.02 (2.83)	91.73 (2.59)	1.71
	Flapped	1,070 (24)	1,080 (37)	1,061 (31)	-19	87.29 (1.83)	87.01 (2.76)	87.57 (2.43)	0.56
	Other	1,090 (20)	1,110 (33)	1,069 (24)	-41	91.17 (1.72)	90.28 (2.57)	92.07 (2.31)	1.79

Note. RTs are given in milliseconds and accuracy rates as percentages. Standard errors are shown in parentheses

Table 4 Mean RTs and accuracy rates for Experiment 2, split by context and stimulus type

	Item type	RT			Accuracy		
		Careful	Casual	Difference	Careful	Casual	Difference
Words	Canonical	915 (19)	916 (22)	1	90.20 (1.15)	91.25 (1.12)	1.05
	Flapped	1014 (27)	986 (24)	-28	72.75 (2.10)	76.61 (2.31)	3.86
	Other	957 (21)	958 (20)	1	92.94 (0.94)	93.14 (0.90)	0.20
Nonwords	Canonical	1,040 (31)	1,034 (30)	-6	90.54 (2.65)	91.28 (2.79)	0.74
	Flapped	1,066 (36)	1,069 (31)	3	87.17 (2.59)	87.14 (2.58)	-0.03
	Other	1,089 (27)	1,084 (27)	-5	91.70 (2.37)	90.83 (2.47)	-0.87

Note. RTs are given in milliseconds and accuracy rates as percentages. Standard errors are shown in parentheses

There was no interaction between variant type and context ($\beta_0 = 2.65, SE = 0.29, z = 9.18; \beta_{\text{Variant} \times \text{Context}} = 0.15, SE = 0.30, z = 0.48, p = .63$) and no main effect of context ($\beta_{\text{Context}} = 0.24, SE = 0.14, z = 1.71, p = .09$). Furthermore, the effect of context on variant type did not change over the course of the experiment ($\beta_0 = 2.65, SE = 0.29, z = 9.19; \beta_{\text{Variant} \times \text{Context} \times \text{Trial}} = -0.29, SE = 0.33, z = -0.86, p = 0.39$). In addition, there was no interaction between context, variant type, and flap-variant frequency ($\beta_0 = 2.51, SE = 0.30, z = 8.25; \beta_{\text{Variant} \times \text{Context} \times \text{Variant Frequency}} = -0.27, SE = 0.62, z = -0.44, p = .66$), showing that the effect of context did not change depending on variant frequency.

Reaction times

As shown in Table 3, the overall RTs for the flapped variants were longer compared with the canonical variants. However, as in Experiment 1, this difference decreased over the course of the experiment. Furthermore, Table 4 shows that the RT difference between the two variants was smaller in the casual-context condition (70 ms) than in the careful-context condition (99 ms). This decrease in the difference between the two variant types is because flapped variants were responded to 28 ms faster in the casual compared with the careful context.

The model in Table 5 shows a significant interaction between variant type and trial. Separate models for each

Table 5 Model parameters for mixed-effects models for accuracy rates and RTs in Experiment 2

	Accuracy				RT			
	β	SE	z value	p value	β	SE	t value	p value
<i>Fixed effects</i>								
Intercept	2.72	0.29	9.31		-0.17	0.04	-4.48	
Variant type	-2.00	0.31	-6.49	<.001	0.28	0.03	10.67	<.001
Context	0.19	0.17	1.11	.27	-0.08	0.03	-2.91	.004
Speed	0.34	0.19	1.77	.08	-1.04	0.05	-22.28	<.001
Trial	0.15	0.10	1.49	.14	-0.02	0.02	-1.34	.18
Previous RT	-	-	-	-	0.04	0.02	2.77	.01
Variant \times Context	0.27	0.34	0.81	.42	-0.09	0.05	-1.85	.06
Variant \times Speed	-0.09	0.38	-0.24	.81	-0.13	0.05	-2.68	.01
Context \times Speed	-0.09	0.37	-0.24	.81	-0.06	0.05	1.40	.16
Variant \times Trial	0.72	0.18	4.01	<.001	-0.06	0.03	-2.36	.02
Variant \times Context \times Speed	0.69	0.68	1.01	.31	0.24	0.12	2.11	.04
<i>Random effects</i>								
	$SD_{\text{participants}}$	SD_{words}			$SD_{\text{participants}}$	SD_{words}		
Intercept	0.71	1.61			0.21	0.13		
Variant type	0.42	1.39			0.00	0.08		
Context	0.00	0.00			0.05	0.00		
Speed	0.00	0.00			0.21	0.16		
Trial	0.18	0.17			0.00	0.03		
Previous RT	-	-			0.06	0.04		
Variant \times Context	0.00	<0.001			0.00	0.08		
Variant \times Speed	0.00	0.70			0.12	0.00		
Context \times Speed	0.96	0.00			0.00	0.01		
Variant \times Trial	0.32	0.00			0.00	0.00		
Variant \times Context \times Speed	0.00	0.00			0.14	0.46		

Note. Trial numbers and RTs were standardized before being entered into the model. SE = standard error, SD = standard deviation

pronunciation variant showed that responses to flapped variants sped up over the course of the experiment ($\beta_{\text{Trial}} = -0.08$, $SE = 0.04$, $t = -1.93$, $p = .05$), whereas there was no effect of trial for canonical variants ($\beta_{\text{Trial}} = -0.03$, $SE = 0.03$, $t = -1.03$, $p = .30$). Separate analyses for each of the two parts of the experiment indicated that the difference in RT between the two variants remained significant over the course of the experiment. However, as indicated by the beta values, the size of the difference was larger in Part 1 ($\beta_{\text{Variant}} = 0.34$, $SE = 0.06$, $t = 5.82$, $p < .001$) compared with Part 2 ($\beta_{\text{Variant}} = 0.24$, $SE = 0.05$, $t = 5.12$, $p < .001$).

In order to examine whether recognition performance for the filler words and nonwords improved, we ran additional models for each stimulus type. These models showed no significant influence of trial number on RT for the filler words ($\beta_{\text{Trial}} = -0.03$, $SE = 0.02$, $t = -1.41$, $p = .16$) or the nonwords with a nasal flap ($\beta_{\text{Trial}} = -0.04$, $SE = 0.04$, $t = -1.10$, $p = .27$). In contrast, responses decreased for the nonwords with a canonical /nt/ sequence ($\beta_{\text{Trial}} = -0.10$, $SE = 0.04$, $t = -2.77$, $p = .01$) and the nonwords with no /nt/ sequence ($\beta_{\text{Trial}} = -0.06$, $SE = 0.03$, $t = -2.19$, $p = .03$).

In order to examine the influence of overall response speed on the processing of variant type and context, we compared the effects of variant type and context across different percentiles of the RT distributions. For this purpose, the distribution of RTs for each condition were divided into 10 percentile bins, and the mean RT was calculated for each bin and each condition. As is shown in in Fig. 4, there is a RT difference between the canonical and flapped variants across all percentiles of the distributions. However, the effect of context appears to emerge only in

the percentiles of 50% and above. More specifically, there is an effect of context for the flapped variants in the slowest 60% of the responses whereas there does not appear to be an effect of context in the fastest 40%. In contrast to the flapped variants, no effect of context is visible in any of the RT percentiles for the canonical variants.

The model in Table 5 shows a three-way interaction between variant type, context, and response speed, suggesting that the effect of context on variant-type processing differs between fast and slow responses. Additional models showed that there was a significant interaction between variant-type and context for the slow responses ($\beta_0 = -0.15$, $SE = 0.06$, $t = -2.36$; $\beta_{\text{Variant} \times \text{Context}} = -0.23$, $SE = 0.11$, $t = -2.12$, $p = .03$), but no such interaction for fast responses ($\beta_0 = 0.21$, $SE = 0.07$, $t = 2.98$; $\beta_{\text{Variant} \times \text{Context}} = 0.16$, $SE = 0.12$, $t = 1.29$, $p = .20$). A model for slow responses to flapped variants showed that words were recognized more quickly in a casual compared with a careful context ($\beta_{\text{Context}} = -0.29$, $SE = 0.08$, $t = -3.64$, $p < .001$). In contrast, a model for slow responses to canonical variants showed no influence of context ($\beta_{\text{Context}} = -0.003$, $SE = 0.07$, $t = -0.04$, $p = .97$). Separate models for each context showed that flapped variants were responded to significantly more slowly compared with canonical variants in both contexts, while the change in the size of the beta values indicates that this difference was larger in the careful than in the casual context (careful context: $\beta_{\text{Variant}} = 0.46$, $SE = 0.06$, $t = 7.61$, $p < .001$; casual context: $\beta_{\text{Variant}} = 0.25$, $SE = 0.07$, $t = 3.58$, $p < .001$).

We explored the three-way interaction between variant type, context, and response speed further by examining

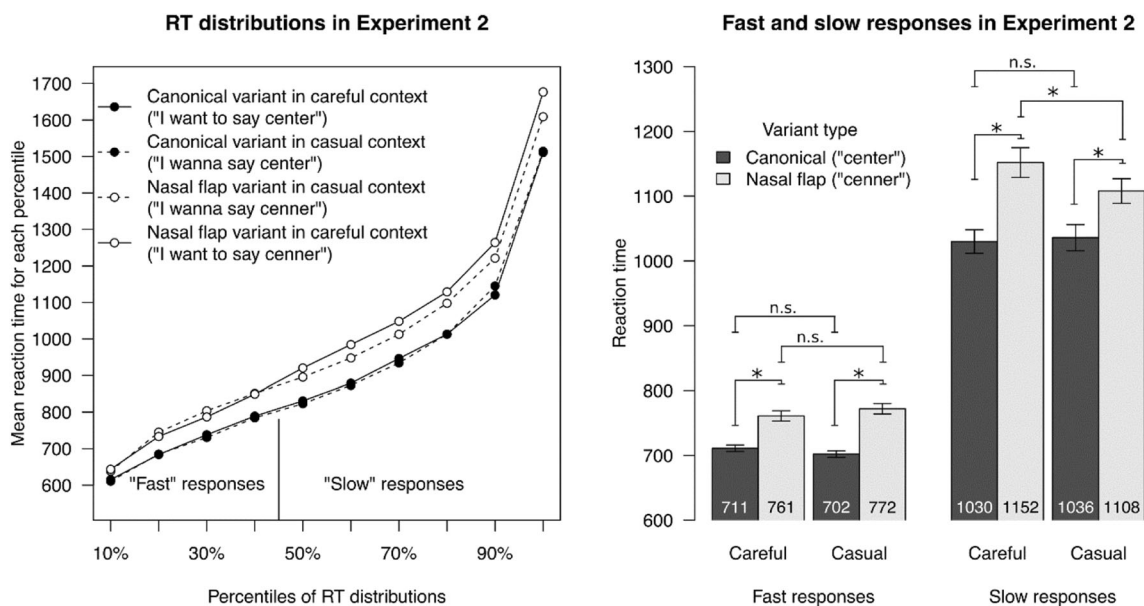


Fig. 4 RT results of Experiment 2. Left: RT distributions for each condition. Right: Mean RTs for each condition, split by fast and slow responses. Error bars indicate one standard error around the mean

whether the size of the interaction effect between context and variant type depends on the overall response speed of participants and words. For this purpose, we fitted a model with only the two-way interaction between variant type and context and included the correlations among the random slopes and intercepts in the random effects structure. Inspection of the random effects for participants and words extracted from this model confirmed that the interaction effect became larger (i.e., more negative, and thereby reducing the positive effect of variant type) as overall RTs increased (see Fig. 5). This finding corroborates that the interaction effect is stronger the slower the overall RTs are. Thus, slow responses appear to be more likely to be influenced by context than fast responses.

In order to examine whether the influence of context on variant-type recognition is influenced by variant frequency, we fit a model including a three-way interaction between variant type, context, and variant frequency. The result showed no significant interaction ($\beta_0 = -0.04$, $SE = 0.10$, $t = -0.37$; $\beta_{\text{Variant} \times \text{Context} \times \text{Variant Frequency}} = 0.14$, $SE = 0.10$, $t = 1.49$, $p = .14$). In order to examine whether the effect of variant frequency might be limited to slow responses to flapped variants, we ran an additional model for only these cases, but found no indication that variant frequency influences the way context is used ($\beta_0 = 0.04$, $SE = 0.08$, $t = 0.57$; $\beta_{\text{Context} \times \text{Variant Frequency}} = 0.07$, $SE = 0.08$, $t = 0.80$, $p = .42$). Furthermore, as in the accuracy analysis, the effect of context on variant type did not change over the course of the experiment ($\beta_0 = -0.02$, $SE = 0.08$, $t = -0.26$; $\beta_{\text{Variant} \times \text{Context} \times \text{Trial}} = 0.05$, $SE = 0.07$, $t = 0.64$, $p = .52$).

Discussion

The results of Experiment 2 replicate the finding from Experiment 1: Listeners’ recognition performance of flapped variants improved over the course of the experiment. However, the main goal of Experiment 2 was to investigate whether phonetic context influences how listeners recognize flapped pronunciation variants. The results show that preceding context can influence variant recognition, but only if the responses are relatively slow. The results show further that the effect of context was limited to flapped variants and that canonical variants did not benefit from a careful carrier phrase, nor were they hampered by a casual carrier phrase. The fact that the context effect emerged only in relatively late responses suggests that context influences word recognition at a relatively late stages during processing. This contrasts with the hypothesis that listeners use context in order to predict upcoming input (e.g., Brunelliere & Soto-Faraco, 2013; Clark, 2013). Instead, phonetic context appears to be used after recognition has already occurred. We will return to this issue in the General Discussion. But first we will conduct a combined analysis of the data from both experiments.

Combined analysis of Experiment 1 and 2

While Experiment 1 and 2 both showed that the recognition of flapped variants improved over the course of the experiment, there were some minor differences in the results of the filler words and nonwords. In order to examine whether these inconsistencies reflect genuine differences between the two

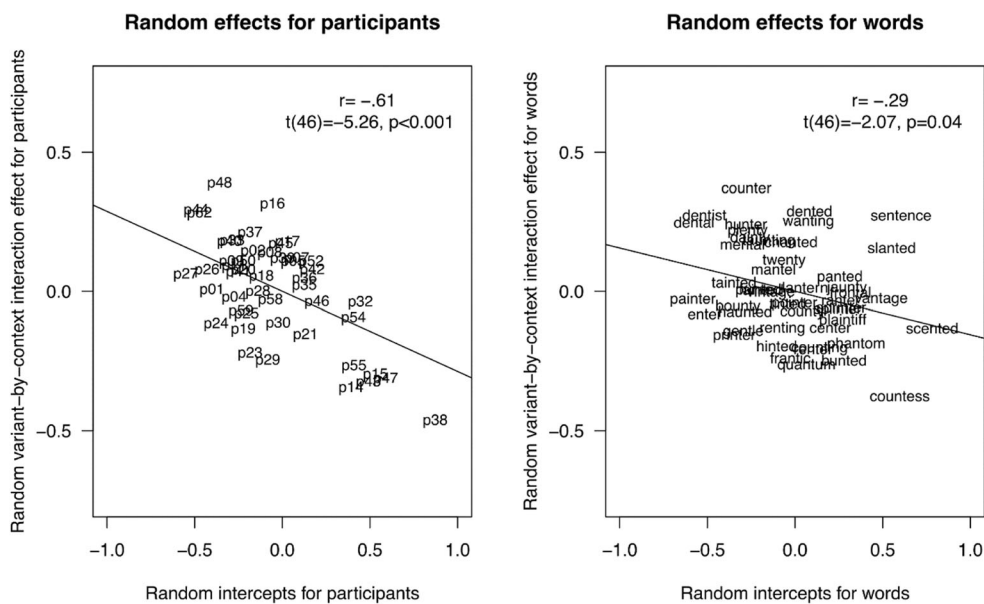


Fig. 5 The correlations between the random intercepts and the random variant-by-context interaction effects for participants (left) and words (right). The diagonal lines are linear regression lines, indicating that the

variant-by-context interaction is stronger for words and participants with overall slow RTs compared with those with overall fast RTs

experiments, we combined both data sets in a single analysis. Furthermore, we examined whether the influence of trial on the variant-type effect differed between Experiments 1 and 2.

Results

We used the same statistical methods as in the previous analyses. The percentage of removed trials in the RT models ranged from 2.76% to 3.37%. The factor *experiment* was entered as a binary variable, with Experiment 1 coded as -0.5 and Experiment 2 coded as 0.5.

Accuracy

Table 6 shows the results of a mixed model based on the combined data sets of Experiments 1 and 2. This model included two-way interactions between experiment and variant type, experiment and trial, as well as a three-way interaction between experiment, variant type, and trial in order to examine whether adaptation to flapped variants differed between experiments. As can be seen in Table 6, there were no effects of experiment, demonstrating that the effects of variant type and trial were identical between the two experiments. Crucially, there is a significant interaction between variant type and trial, showing that listeners adapted to flapped variants in both experiments.

Additional models showed that flapped variants were responded to less accurately compared with canonical variants in both the first part ($\beta_{\text{Variant}} = -2.13, SE = 0.28, z = -7.58, p <$

.001) and the second part ($\beta_{\text{Variant}} = -1.36, SE = 0.25, z = -5.45, p < .001$) of the two experiments. But, as indicated by the beta values, this effect decreased from the first to the second part. This finding is further supported by a significant increase in accuracy across trials for the flapped variants ($\beta_{\text{Trial}} = 0.33, SE = 0.07, z = 4.81, p < .001$), whereas accuracy for canonical variants decreased slightly ($\beta_{\text{Trial}} = -0.18, SE = 0.09, z = -2.04, p = .04$).

We also examined the filler words as well as the nonwords. For none of these stimulus types did we find an interaction between trial and experiment (all z s ≤ 1.33 and all p s $\geq .18$). Models without experiment as a predictor variable showed a significant decrease in accuracy across trials for the filler words ($\beta_{\text{Trial}} = -0.20, SE = 0.06, z = -3.43, p = .29$). There was no effect of trial for the nonwords with a canonical /nt/ sequence ($\beta_{\text{Trial}} = 0.10, SE = 0.13, z = 0.79, p = .43$) or for nonwords with a nasal flap ($\beta_{\text{Trial}} = 0.10, SE = 0.08, z = 1.25, p = .21$). For the nonwords without /nt/ sequence, there was a marginally significant increase in accuracy ($\beta_{\text{Trial}} = 0.13, SE = 0.06, z = 1.97, p = .05$).

There was no four-way interaction between variant type, variant frequency, trial, and experiment ($\beta_0 = 2.46, SE = 0.23, z = 10.85; \beta_{\text{Variant} \times \text{Trial} \times \text{Variant Frequency} \times \text{Experiment}} = -0.48, SE = 0.46, z = -1.06, p = .29$). We also found no three-way interaction between variant type, trial, and variant frequency when collapsing the data across both experiments ($\beta_0 = 2.38, SE = 0.22, z = 11.07; \beta_{\text{Variant} \times \text{Trial} \times \text{Variant Frequency}} = -0.09, SE = 0.19, z = -0.45, p = .65$). These results suggest that in neither experiment did variant frequency influence adaptation to flapped variants.

Table 6 Model parameters for mixed-effects models for RTs and accuracy rates for the combined analysis of Experiment 1 and 2

	Accuracy				RT			
	β	<i>SE</i>	<i>z</i> value	<i>p</i> value	β	<i>SE</i>	<i>t</i> value	<i>p</i> value
<i>Fixed effects</i>								
Intercept	2.51	0.22	11.36		-0.02	0.06	-0.36	
Variant type	-1.81	0.23	-7.91	<.001	0.33	0.04	8.63	<.001
Trial	0.09	0.06	1.62	.11	-0.05	0.02	-2.79	.01
Experiment	-0.02	0.18	-0.12	.90	0.002	0.08	0.02	.98
Previous RT	-	-	-	-	0.09	0.02	5.81	<.001
Variant \times Trial	0.51	0.11	4.73	<.001	-0.08	0.02	-3.24	.001
Variant \times Experiment	0.01	0.24	0.04	.97	-0.02	0.05	-0.36	.72
Trial \times Experiment	0.12	0.14	0.84	.40	-0.02	0.03	-0.46	.64
Variant \times Trial \times Exp.	0.29	0.22	1.34	.18	-0.01	0.04	-0.21	.84
<i>Random effects</i>								
	<i>SD</i> _{participants}	<i>SD</i> _{words}			<i>SD</i> _{participants}	<i>SD</i> _{words}		
Intercept	0.71	1.36			0.39	0.34		
Variant type	0.48	1.30			0.13	0.21		
Trial	0.23	0.00			0.14	0.06		
Experiment	-	0.36			-	0.03		
Previous RT	-	-			0.10	0.07		
Variant \times Trial	0.23	0.00			0.00	0.08		
Variant \times Experiment	-	0.59			-	0.00		
Trial \times Experiment	-	0.50			-	0.03		
Variant \times Trial \times Exp.	-	<0.001			-	<0.001		

Note. Trial numbers and RTs were standardized before being entered into the model. *SE* = standard error; *SD* = standard deviation, Exp. = experiment

Reaction times

The model shown in Table 6 indicates that there were no effects of Experiment, suggesting that the effects of variant type and trial were identical between the two experiments. Crucially, there was a significant interaction between variant type and trial, showing that listeners' ability to recognize flapped variants improved in both experiments. Additional models showed that flapped variants were responded to more slowly compared with canonical variants in both the first part ($\beta_{\text{Variant}} = 0.35$, $SE = 0.04$, $t = 7.88$, $p < .001$) and the second part ($\beta_{\text{Variant}} = 0.26$, $SE = 0.04$, $t = 6.18$, $p < .001$) of the two experiments. But, as indicated by the beta values, this effect was decreased over the course of the experiment. This finding is further supported by a significant decrease in RT across trials for the flapped variants ($\beta_{\text{Trial}} = -0.09$, $SE = 0.03$, $t = -3.48$, $p < .001$), but not for the canonical variants ($\beta_{\text{Trial}} = -0.02$, $SE = 0.02$, $t = -1.14$, $p = .25$).

We also examined the filler words as well as the nonwords. For none of these stimulus types did we find an interaction between trial and experiment (all $|ts| \leq 1.81$ and all $ps \geq .07$). Models without experiment as a predictor variable showed no significant effect of trial for the filler words ($\beta_{\text{Trial}} = -0.01$, $SE = 0.01$, $t = -0.89$, $p = .37$). In contrast, RTs decreased across trials for the nonwords with a canonical /nt/ sequence ($\beta_{\text{Trial}} = -0.09$, $SE = 0.02$, $t = -3.84$, $p < .001$), the nonwords with a nasal flap ($\beta_{\text{Trial}} = -0.06$, $SE = 0.02$, $t = -3.33$, $p < .001$), and the nonwords without /nt/ sequence ($\beta_{\text{Trial}} = -0.07$, $SE = 0.02$, $t = -3.38$, $p = .001$). As in the accuracy analysis, the RT analysis showed no four-way interaction between variant type, variant frequency, trial, and experiment ($\beta_0 = 0.01$, $SE = 0.09$, $t = 0.10$; $\beta_{\text{Variant} \times \text{Trial} \times \text{Variant Frequency} \times \text{Experiment}} = 0.05$, $SE = 0.06$, $t = 0.80$, $p = .42$). Furthermore, there was no three-way interaction between variant type, trial, and variant frequency when collapsing the data across both experiments ($\beta_0 = 0.01$, $SE = 0.09$, $t = 0.07$; $\beta_{\text{Variant} \times \text{Trial} \times \text{Variant Frequency}} = -0.04$, $SE = 0.03$, $t = -1.16$, $p = .25$). These results mirror the accuracy results and suggest that in neither experiment did variant frequency influence adaptation to flapped variants.

Discussion

The results from the combined analysis show no difference in adaptation across the two experiments and support the main findings from the analyses of the previous experiments: Listeners' ability to recognize flapped variants improves over the course of the experiment, whereas listeners' ability to recognize canonical words (with or without /nt/ sequence) does not improve. This shows that the processing disadvantage for flapped variants decreases over time. One possible alternative explanation is that this performance improvement is due to some general adaptation to the lexical decision task, rather than reflective of adaptation that is specific to flapped variants.

For example, listeners may have gotten used to the task and therefore were able to respond more quickly. While it is true that RTs to nonwords sped up somewhat, we do not find any evidence suggesting that performance for the filler or canonical words improved. This suggests that listeners became more familiar with nonwords. Participants are unlikely to encounter nonwords outside of psycholinguistic experiments and so they do not know what nonwords are and how to categorize them when entering the lab. In order to perform the lexical decision task, participants need to learn "nonword" as a new response category, meaning that participants have to learn what items are like that belong into this novel category. In contrast, participants do not have to learn a new response category for "words" because participants already know what words are. This learning process for the nonword response category is probably the reason for the decrease in nonword RTs over the course of the experiment. Ideally, this task-specific learning effect is captured by the practice trials. However, in our study the number of practice nonwords was relatively small (only six) and thus it seems that this effect spilled over into the main experiment. It is likely that with a larger number of practice trials the improvement effect for nonwords would be smaller or even absent. But since RTs for filler and canonical words did not change over the course of the experiment, the performance improvement for flapped words cannot be due to general adaptation to the task and instead shows that listeners learned something that is specific to flapped pronunciation variants.

Another possibility is that listeners changed their decision criterion for distinguishing between words and nonwords. If this were true, we would expect to see more incorrect responses to nonwords because listeners should be more likely to accidentally accept nonwords as words. However, accuracy for nonwords remained constant across the experiment, suggesting that the decision criterion did not change. A related possibility is that faster responses are due to a speed–accuracy trade-off reflecting the fact that listeners sacrificed accuracy in order to speed up their responses. However, if this were true, we would expect to not only see faster responses to flapped variants but also to the filler and canonical words, which was not the case. Furthermore, a speed–accuracy trade-off would be inconsistent with the increase in accuracy rates for the flapped variants.

In sum, none of these task-related explanations can account for the present pattern of results. Instead, the results show (1) that listeners responded more quickly and accurately to flapped pronunciation variants, (2) that listeners became more familiar with the nonwords, and (3) that accuracy to canonical variants (both with and without /nt/) decreased slightly. The last finding is somewhat peculiar. One possibility is that there was a small speed–accuracy trade-off that did not reach significance in the RTs, but did become significant in the accuracy rates. However, if there was such a trade-off, it was

certainly not big enough to explain the significant performance improvement in RT *and* accuracy for flapped variants. Another possibility is that as listeners adapted to flapped variants, canonically produced word forms started to sound more artificial, which may have resulted in more accidental non-word responses. The most interesting result, however, is the performance improvement with respect to the flapped variants, which we will discuss in the following section.

General discussion

The goal of the present study was to examine the role of listener flexibility in the recognition of allophonic variants that are produced with nasal flaps. Previous research suggested that flapped variants are processed more slowly and less accurately compared with canonical variants. In the present study, we investigated whether exposure to flapped variants and phonetic context can decrease this reported processing disadvantage for flapped variants. Experiment 1 showed that the processing disadvantage for flapped variants compared with canonical variants decreased over the course of the experiment. Whereas flapped variants were responded to more slowly and less accurately than canonical variants in the beginning of the experiment, this processing disadvantage decreased considerably toward the end of the experiment. Experiment 2 replicated this effect and showed further that flapped variants are recognized more quickly if they are embedded in a casually produced sentence frame than when they occur in a carefully produced context. Interestingly, the effect of context was limited to responses that were relatively slow. In the following, we will discuss how the performance improvements due to increased exposure and matching phonetic context could be explained.

Examination of the canonical variants and filler words showed that the performance improvement for flapped variants was not due to a general adaptation to the task, but that listeners learned something that was specific to the flapped pronunciation variants. What might it be that the listener learned about flapped variants? According to the episodic view, listeners store individual representations of different phonological variants in the mental lexicon (e.g., Goldinger, 1998). Thus, both the canonical variant *center* and the flapped variant *cenner* have their own lexical representation in the listeners' memory. Evidence for this claim comes from effects of variant frequency on the recognition and production of pronunciation variants (e.g., Bürki, Ernestus, & Frauenfelder, 2010; Pitt, Dilley, & Tat, 2011; Ranbom & Connine, 2007). These studies have shown that the frequency with which reduced variants occur have a large influence on how quickly they can be recognized. However, the exemplar view would have to make additional assumptions in order to explain how exposure to the flapped variant of one word can

facilitate the recognition of the flapped variant of another word. Furthermore, the fact that we did not find an interaction between trial or context and variant frequency is inconsistent with the episodic view. Adaptation ought to be stronger for less frequent variants as these should benefit more from additional information. The absence of an influence of variant frequency is more consistent with an account which operates independently of variant frequency.

A different explanation might be offered by hybrid theories that assume that listeners do not only store linguistic information when learning a new word but also social information associated with the learning situation. According to the dual-route approach to speech perception, stored knowledge about word forms interacts with social knowledge (Sumner et al., 2014). This account could explain the decrease in the processing disadvantage for flapped variants by assuming that hearing flapped variants activates sociophonetic knowledge about flapping. Once these social representations are sufficiently activated, they could facilitate the recognition of flapped variants.

In contrast to episodic and hybrid theories, abstractionist theories assume that listeners store only a single abstract phonological representation for each word. Instead of assuming that listeners learn new lexical representations when they encounter novel phonological variants, this approach focuses on the learning of sublexical representations. According to this view, listeners are constantly learning about phonological categories while being exposed to speech. Evidence for this approach comes from perceptual learning studies that demonstrated that listeners' phonetic categorizations change after being exposed to ambiguous speech sounds presented in biasing lexical contexts (e.g., Clarke-Davidson et al., 2008; Mitterer & Reinisch, 2013; Norris et al., 2003; Samuel & Kraljic, 2009). According to this approach, listeners in our experiment may have learned to map the nasal flap onto the same phonemic category as the /nt/ sequence. Note that the listeners in our experiment were exposed to both canonical and flapped variants, and that recognition performance for canonical variants did not change over the course of the experiment. This is consistent with the finding that perceptual learning leads to an expansion of phonological categories rather than a replacement of the category content (e.g., McQueen, Cutler, & Norris, 2006). Our results are more compatible with hybrid and abstractionist views of the mental lexicon than with the episodic view. In particular, the fact that listeners adapted to nasal flapping across different word types is difficult to reconcile with the episodic account.

The findings of Experiment 1 were replicated and extended by Experiment 2, which showed that the recognition of flapped variants is also improved by phonetic context. Flapped variants were responded to more quickly if they were preceded by a casually produced carrier sentence (*I wanna say . . .*) compared with when they were preceded by a carefully

produced carrier sentence (*I want to say . . .*). Interestingly, we did not find a context effect for canonical variants, showing that—in contrast to flapped variants—canonical variants neither benefitted from matching phonetic context nor was their recognition hampered by mismatching context. This finding is consistent with previous work on the influence of phonetic context on the recognition of regional accents. In Brunelliere and Soto-Faraco's (2013) ERP study, a preceding nonnative context facilitated the recognition of nonnative target words compared with a preceding native context. In contrast, a preceding native context did not facilitate the recognition of native target words compared with a preceding nonnative context. This finding is at odds with a purely episodic view of speech recognition and is more consistent with abstractionist and hybrid accounts that assume that the canonical form has a privileged status compared with noncanonical forms (e.g., Ranbom & Connine, 2007; Sumner et al., 2014).

The dual-route approach to speech perception (Sumner et al., 2014) could explain the context effect in the same way that it explains the effect of increased exposure. In addition to activating linguistic word-form representations, the casually produced context also activates sociophonetic representations that are associated with nasal flapping. These representations then facilitate the recognition of flapped pronunciation variants. The build-up of activation of sociophonetic representations and the subsequent facilitation of variant recognition may require some time, which would explain why the context effect occurs only for relatively slow responses.

Abstractionist theories could explain the context effect in terms of context-induced perceptual learning. The basic principle that underlies perceptual learning is that listeners are able to use some cue to tell them what a stimulus is *supposed to be* (Cutler, 2012). Listeners may have used the word *wanna* in the casual carrier phrase *I wanna say . . .* in order to tell them that the nasal flap is supposed to be mapped onto the /nt/ sequence. This inference may have occurred online, requiring additional processing time, which would be consistent with the fact that the context effect emerged only in relatively late RTs.

The finding that the context effect in Experiment 2 occurred relatively late is interesting because it suggests that phonetic context is not used in order to predict upcoming phonetic input (e.g., Clark, 2013; DeLong & Kutas, 2008). Instead phonetic context appears to exert its influence rather late, that is, after recognition has already taken place. There are several possibilities why listeners did not use the context for prediction. One possibility is that listeners may require more time in order to generate predictions based contextual information. Similarly, it may be the case that listeners require more phonetic information in order to make a prediction. Had the preceding context be longer than the phrase *I wanna say . . .*, listeners might have used the additional time and/or the additional phonetic information in order to predict nasal

flapping. Another possibility is that listeners did not use the context for prediction given that half of the time the speaking style of the carrier phrase did not match the pronunciation of the target word. In a recent study, Fine et al. (2013) demonstrated that participants rapidly adjust their expectations of upcoming syntactic structures depending on the predictability of the syntactic structures in the experiment. It is possible that in our experiment, the absence of prediction is due to adaptation to the trials with mismatching pairings of context and target word. We can therefore not rule out the possibility that listeners can under different circumstances use phonetic context in order to predict the upcoming speaking style. Further research could explore this issue by providing participants with a more immersive and consistent phonetic context. Another possible reason why listeners did not use contextual information for prediction in our experiment is related to the fact that the durations of the canonical and flapped variants were equalized. As was mentioned earlier, the differences in length between the original and modified recordings were barely detectable, and both versions sounded very natural. Nevertheless, in principle it is still possible that the change in duration introduced an additional type of mismatch which could interfere with expectations (similar to the mismatch between a casually produced phonetic context and a canonically produced target word). This issue could be addressed in future studies by replicating the current finding with unmodified stimuli.

The notion that listeners use contextual information at later processing stages is consistent with previous work that has shown that the speed with which listeners respond to words influences the degree to which indexical information is processed (McLennan & Luce, 2005). Indexical information (i.e., variability in talker identity and speaking rate) appears to be stored only if processing is relatively hard which results in slow responses (but see also Papesh, Goldinger, & Hout, 2016). Similarly, the influence of phonetic context may also require time to affect word recognition because it takes listeners time to integrate the target word with the preceding input. Processing speech relatively slowly therefore allows for more time to integrate the preceding context with a given word which allows listeners to adapt the way they recognize words by making use of contextual information.

Taking more time during speech perception may thus be a useful strategy when listening to casual speech because it allows listeners to exploit valuable information provided by the phonetic context. This idea is consistent with the strategy of *delayed commitment* (Marr, 1982; see also Cluff & Luce, 1990; Klatt, 1989). According to this strategy, listeners withhold all-or-none decisions as long as possible in order to avoid having to backtrack in the case of degraded or ambiguous input. Although delaying commitment to a particular word candidate (or a lexical decision) may lead to longer response times, such a strategy also minimizes a costlier reinvestment

of resources if the initial decision turns out to be wrong. This strategy also allows listeners more time to integrate a given item with preceding acoustic-phonetic as well as the pragmatic or sentential context. Previous studies have suggested that delayed commitment is a viable strategy for dealing with temporary lexical ambiguity in spoken language due to lexical embedding (Luce & Cluff, 1998). This strategy may also be useful for recognizing reduced speech because it allows the listeners to take more time in order to consider information from the preceding and following phonetic context.

In conclusion, the results of the present study show that flexibility plays an important role in how listeners cope with allophonic variation. Given enough time to adapt and the appropriate kind of context, listeners' ability to recognize flapped pronunciation variants improves considerably. Our results are therefore consistent with previous studies that suggest that the processing costs for reduced variants may be smaller than previously suggested (e.g., Bürki, Viebahn, Racine, Mabut, & Spinelli, 2017; Sumner, 2013). When not considering the listener's capacity to adapt, we may overestimate the processing advantage for canonical forms and underestimate the listener's ability to cope with variable linguistic input.

Acknowledgements We thank Jim Sawusch and Micah Geer for comments on earlier versions of this manuscript and Ashley Williams, Hital Patel, and Katie Simon for assistance in testing participants.

Appendix

Words with /nt/ sequence: bounty, center, counter, counting, county, dainty, dental, enter, frontal, gentle, hunter, mantel, mental, painter, plaintiff, plenty, renting, taunting, vantage, vantage, bunted, chanted, countless, dented, dentist, fainted, frantic, haunted, hinted, jaunty, lantern, panted, phantom, pointer, printer, punted, quantum, ranter, renter, scented, sentence, slanted, splinter, sprinter, tainted, tinted, twenty, wanting

Nonwords with /nt/ sequence (in Klattese): bUntu, sintL, kYntL, kYntxm, kUnto, dYnte, dOntX, YntL, fr@ntX, JRntX, huntL, mUntX, mintxn, pontL, plcntxs, plWnto, rentxm, tontxm, vWntxS, vOntxz, prYntxd, t@ntxd, krIntxs, bentxd, gantxst, CYntx, CWntxk, J^ntxd, sOntxd, zonti, fUntRn, suntxd, vRntxm, SintX, h@ntX, nintxd, mYntxm, sWntX ContX, rUntxd, wIntxns, Yintxd, slRntR, skrentR, pRntxd, kUntxd, bInti, d@ntIG

Filler words: alcove, alpine, anger, aspect, Aspen, basket, biscuit, campus, casket, childish, compact, compass, compost, compound, dampen, discard, discuss, elder, fiscal, folder, golden, hectic, helper, holding, impulse, jumper, mascot, olden, pamper, pompous, prospect, prosper, pulpit, rascal, scalpel, seldom, shoulder, smoulder, suspect, temper, trespass,

trumpet, Tuesday, vulgar, welcome, whimper, whisper, wisdom

Nonwords without /nt/ sequence (in Klattese): ilkIv, Ilpin, OGx, Espekt, espxn, paskxt, tWskxt, bYmpxs, dcskxt, g@ldxS, COmpEkt, Jompxs, fUmp@st, tRmpYnd, vimpxn, seskRd, h@skus, ildX, nusxkl, mRldX, lildxn, rIktkk, w@lpX, yaldIG, Wmpels, YmpX, pr^skit, Yldxn, tmpX, krOmpxs, bospikt, dUspX, gruspxt, CRskxl, SIlpxl, zeldxm, f@ldX, tildX, SEspxkt, vempX, l@spEs, rampxt, wUsdW, yYlgX, smOlkxm, trimpX, pUspX, tEsdxm

Practice words: vector, present, cancel, salary, purple, buckle

Practice nonwords: trUflo, kOfet, flUSu, flofL, klefxm, SrifL

References

- Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. doi:[https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1)
- Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, 28(4), 525–560. doi:<https://doi.org/10.1080/01690965.2012.658072>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Baayen, R. H., & Milin, P. (2010). Analyzing Reaction Times. *International Journal of Psychological Research*, 3(2), 12–28.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. doi:<https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D. M. (2005). Fitting linear mixed models in R: Using the lme4 package. *R News: The Newsletter of the R Project*, 5(1), 27–30.
- Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.1.35) [Computer software]. Retrieved from <http://www.praat.org>
- Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, 27(4), 539–571. doi:<https://doi.org/10.1080/01690965.2011.555268>
- Brunelliere, A., & Soto-Faraco, S. (2013). The speakers' accent shapes the listeners' phonological predictions during speech perception. *Brain and Language*, 125(1), 82–93. doi:<https://doi.org/10.1016/j.bandl.2013.01.007>
- Bürki, A., Ernestus, M., & Frauenfelder, U. H. (2010). Is there only one “fenêtre” in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words. *Journal of Memory and Language*, 62(4), 421–437. doi:<https://doi.org/10.1016/j.jml.2010.01.002>
- Bürki, A., Viebahn, M. C., Racine, I., Mabut, C., & Spinelli, E. (2017). Intrinsic advantage for canonical forms in spoken word recognition: Myth or reality? *Language, Cognition and Neuroscience*, 33(4), 494–511. doi:<https://doi.org/10.1080/23273798.2017.1388412>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204. doi:<https://doi.org/10.1017/S0140525X12000477>
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of*

- America, 116(6), 3647–3658. doi:<https://doi.org/10.1121/1.1815131>
- Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics*, 70(4), 604–618. doi:<https://doi.org/10.3758/PP.70.4.604>
- Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two syllable words: Retroactive effects on multiple activation. *The Journal of the Acoustical Society of America*, 87(S1), S125–S126. doi:<https://doi.org/10.1121/1.2027912>
- Coenen, E., Zwitserlood, P., & Boelte, J. (2001). Variation and assimilation in German: Consequences for lexical access and representation. *Language and Cognitive Processes*, 16(5/6), 535–564.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25, 257–271.
- Connine, C. M., & Darnieder, L. M. (2009). Perceptual learning of co-articulation in speech. *Journal of Memory and Language*, 61(3), 412–422. doi:<https://doi.org/10.1016/j.jml.2009.07.003>
- Cutler, A. (2012). Native listening. Cambridge: MIT Press.
- Cutler, A., Treiman, R., & van Ooijen, B. (1998). Orthografiekonkistentie epektis in foneme detektion? *Proceedings of the Fifth International Conference on Spoken Language Processing*, 6, 2783–2786.
- Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 498–513. doi:<https://doi.org/10.1037/0278-7393.30.2.498>
- DeLong, K. A., & Kutas, M. (2008). Brainwave studies of contextual preactivation of lexical word forms. *Psychophysiology*, 45(1), S9–S10.
- R Development Core Team. (2007). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- Ernestus, M., Baayen, R. H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81(1/3), 162–173. doi:<https://doi.org/10.1006/brln.2001.2514>
- Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39(3), 253–260. doi:[https://doi.org/10.1016/S0095-4470\(11\)00055-6](https://doi.org/10.1016/S0095-4470(11)00055-6)
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLOS ONE*, 8(10), e77661. doi:<https://doi.org/10.1371/journal.pone.0077661>
- Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 380–396. doi:<https://doi.org/10.1037/0096-1523.24.2.380>
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.
- Grainger, J., Muneaux, M., Farioli, & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology*, 58(6), 981–998. doi:<https://doi.org/10.1080/02724980443000386>
- Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, 24(4), 878–887.
- Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, 31(1), 19–31. doi:<https://doi.org/10.1080/23273798.2015.1072223>
- Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, 98(Suppl. C), 1–11. doi:<https://doi.org/10.1016/j.jml.2017.09.002>
- Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous speech: Data and analysis* (pp. 29–54). Tokyo: The National International Institute for Japanese Language.
- Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, 32(1), 37–55. doi:<https://doi.org/10.1023/A:1021933015362>
- Klatt, D. H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge: MIT Press.
- Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, 107(1), 54–81. doi:<https://doi.org/10.1016/j.cognition.2007.07.013>
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.
- Ladefoged, P. (2000). *A course in phonetics* (5th). San Diego: Harcourt, Brace, and Jovanovich.
- Luce, P. A., & Cluff, M. S. (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. *Perception & Psychophysics*, 60(3), 484–490. doi:<https://doi.org/10.3758/BF03206868>
- Luce, P. A., & McLennan, C. T. (2005). Spoken word recognition: The challenge of variation. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 590–609). Hoboken: Blackwell. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/9780470757024.ch24.summary>
- Luce, P. A., & Pisoni, D. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1), 1–36. doi:<https://doi.org/10.1097/00003446-199802000-00001>
- Magnuson, J. S., Tanenhaus, M. K., & Aslin, R. N. (2008). Immediate effects of form-class constraints on spoken word recognition. *Cognition*, 108(3), 866–873. doi:<https://doi.org/10.1016/j.cognition.2008.06.005>
- Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.
- McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 306–321. doi:<https://doi.org/10.1037/0278-7393.31.2.306>
- McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(4), 539–553. doi:<https://doi.org/10.1037/0278-7393.29.4.539>
- McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30(6), 1113–1126. doi:https://doi.org/10.1207/s15516709cog0000_79
- Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, 34(1), 73–103. doi:<https://doi.org/10.1016/j.wocn.2005.03.003>
- Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 244–263. doi:<https://doi.org/10.1037/a0012730>
- Mitterer, H., & Reinisch, E. (2013). No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language*, 69(4), 527–545. doi:<https://doi.org/10.1016/j.jml.2013.07.002>
- Mitterer, H., & Reinisch, E. (2015). Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language*, 85, 116–134. doi:<https://doi.org/10.1016/j.jml.2015.08.005>

- Mouliès, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9, 453–467.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238. doi:[https://doi.org/10.1016/S0010-0285\(03\)00006-9](https://doi.org/10.1016/S0010-0285(03)00006-9)
- Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2016). Eye movements reveal fast, voice-specific priming. *Journal of Experimental Psychology: General*, 145(3), 314–337.
- Patterson, D., LoCasto, P. C., & Connine, C. M. (2003). Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica*, 60(1), 45–69. doi:<https://doi.org/10.1159/000070453>
- Perre, L., Pattamadilok, C., Montant, M., & Ziegler, J. C. (2009). Orthographic effects in spoken language: On-line activation or phonological restructuring? *Brain Research*, 1275, 73–80. doi:<https://doi.org/10.1016/j.brainres.2009.04.018>
- Pitt, M. A. (2009). The strength and time course of lexical activation of pronunciation variants. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 896–910. doi:<https://doi.org/10.1037/a0013160>
- Pitt, M. A., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics*, 39(3), 304–311. doi:<https://doi.org/10.1016/j.wocn.2010.07.004>
- Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, 57(2), 273–298. doi:<https://doi.org/10.1016/j.jml.2007.04.001>
- Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, 71(6), 1207–1218. doi:<https://doi.org/10.3758/APP.71.6.1207>
- Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, 5(6), 546–554. doi:<https://doi.org/10.1037/0278-7393.5.6.546>
- Sumner, M. (2013). A phonetic explanation of pronunciation variant effects. *The Journal of the Acoustical Society of America*, 134(1), EL26–EL32. doi:<https://doi.org/10.1121/1.4807432>
- Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology*, 4. doi:<https://doi.org/10.3389/fpsyg.2013.01015>
- Taft, M., Castles, A., Davis, C., Lazendic, G., & Nguyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language*, 58(2), 366–379. doi:<https://doi.org/10.1016/j.jml.2007.11.002>
- Tuinman, A., Mitterer, H., & Cutler, A. (2014). Use of syntax in perceptual compensation for phonological reduction. *Language and Speech*, 57(1), 68–85. doi:<https://doi.org/10.1177/0023830913479106>
- van Berkum, J., Brown, C., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443–467. doi:<https://doi.org/10.1037/0278-7393.31.3.443>
- van de Ven, M., Tucker, B. V., & Ernestus, M. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory & Cognition*, 39(7), 1301–1316. doi:<https://doi.org/10.3758/s13421-011-0103-2>
- Viebahn, M. C., Ernestus, M., & McQueen, J. M. (2015). Syntactic predictability in the recognition of carefully and casually produced speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(6), 1684–1702.
- Viebahn, M. C., Ernestus, M., & McQueen, J. M. (2017). Speaking style influences the brain's electrophysiological response to grammatical errors in speech comprehension. *Journal of Cognitive Neuroscience*, 29(7), 1132–1146. doi:https://doi.org/10.1162/jocn_a_01095
- Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, 75(3), 537–556.