Article

Language *and* Speech

Language *and* Speech
2014, Vol. 57(1) 68–85

DOI: 10.1177/0023830913479106
las.sagepub.com

Use of Syntax in Perceptual Compensation for Phonological Reduction

Annelie Tuinman
Max Planck Institute for Psycholinguistics, The Netherlands

Holger Mitterer
Max Planck Institute for Psycholinguistics, The Netherlands

Anne Cutler
Max Planck Institute for Psycholinguistics, The Netherlands
MARCS Institute, University of Western Sydney, Australia
Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands

Abstract

Listeners resolve ambiguity in speech by consulting context. Extensive research on this issue has largely relied on continua of sounds constructed to vary incrementally between two phonemic endpoints. In this study we presented listeners instead with phonetic ambiguity of a kind with which they have natural experience: varying degrees of word-final /t/-reduction. In two experiments, Dutch listeners decided whether or not the verb in a sentence such as *Maar zij ren(t) soms* 'But she sometimes run(s)' ended in /t/. In Dutch, presence versus absence of final /t/ distinguishes third- from first-person singular present-tense verbs. Acoustic evidence for /t/ varied from clear to absent, and immediately preceding phonetic context was consistent with more versus less likely deletion of /t/. In both experiments, listeners reported more /t/s in sentences in which /t/ would be syntactically correct. In Experiment 1, the disambiguating syntactic information preceded the target verb, as above, while in Experiment 2, it followed the verb. The syntactic bias was greater for fast than for slow responses in Experiment 1, but no such difference appeared in Experiment 2. We conclude that syntactic information does not directly influence pre-lexical processing, but is called upon in making phoneme decisions.

Keywords

phonological reduction, syntactic context, phonetic ambiguity, decision bias

Corresponding author:
Anne Cutler, The MARCS Institute, University of Western Sydney, Locked Bag 1797, Penrith South, NSW 2751, Australia.
Email: A.Cutler@uws.edu.au

1 Introduction

What we expect to see or hear influences what we think that we perceive. The sequences A 13 C and 12 13 14 look respectively like three consecutive letters and three consecutive numbers. However, the second item in each sequence is identical; '13' is interpreted as B when flanked by letters but as 13 when flanked by numbers, and this difference is due to the context.

In the case of speech perception, explaining such effects of context on interpretation has generated a substantial literature, principally concerning lexical influences on the identification of phonemes. Ganong (1980) developed the classic paradigm for investigating this: listeners decide about ambiguous stimuli on a phonetic continuum, with stimuli from the continuum being presented in syllables for which only one endpoint forms a word. Thus Ganong tested a voicing continuum from [d] to [t] syllable-initially in *dice–tice* or *dype–type*. He found that listeners tended to label the perceptually most ambiguous stimuli such that the utterance formed an existing word (as [d] before *-ice*, but as [t] before *-ype*). The effect has been replicated in word-medial position (e.g., [d], [g] in *cradle–cragle* and *badel–bagel*: Connine, 1990) and word-final position (e.g., [s], [ʃ] in *kiss–kish* and *fiss–fish*: McQueen, 1991).

Models of spoken-word recognition, including TRACE (McClelland & Elman, 1986), the Distributed Cohort Model (Gaskell & Marslen-Wilson, 1997), the Race model (Cutler, Mehler, Norris, & Segui, 1987), and Shortlist A and B (Norris, 1994; Norris & McQueen, 2008) agree in distinguishing pre-lexical representations of speech input from stored representations of words, but disagree about the nature and possible directions of the information flow between the pre-lexical and lexical processing levels. In TRACE, information can flow from the lexicon back to the pre-lexical level and directly influence perception. In autonomous models such as Shortlist or Race, in contrast, such reverse information flow is impossible. The models thus differ in how such lexical effects are explained: as direct top-down influence of the lexicon on pre-lexical processing (e.g., TRACE), or in terms of integration of independent information sources (e.g., the Merge model; Norris, McQueen, & Cutler, 2000). In Merge, information flow is strictly bottom-up, and phonemic decisions are made by dedicated decision units that continuously receive and merge input from pre-lexical and lexical levels of processing. Thus both information sources influence the phoneme decisions, without feedback of lexical information to pre-lexical processing.

Sentential context, too, affects the interpretation of ambiguous word tokens. Listeners are more likely to identify a word as the alternative that makes most sense in the sentence they have heard (Borsky, Shapiro, & Tuller, 2000; Borsky, Tuller, & Shapiro, 1998; Connine, 1987; Connine, Blasko, & Hall, 1991; Miller, Green, & Schermer, 1984; Rohde & Ettlinger, 2012). In Borsky et al.'s (1998) study, for example, listeners heard stimuli from a *goat–coat* continuum embedded in sentences where a verb such as *milk* or *dry-clean* biased interpretation to either *goat* or *coat*. At the offset of the ambiguous *goat–coat* stimulus, listeners saw a probe word (*goat* or *coat*) and decided whether the visual word matched what they had just heard. A sentential context effect appeared; listeners gave more *goat* identifications after *milk* and more *coat* responses after *dry-clean*. Again, competing accounts ascribe such sentential context effects to feedback (e.g., Borsky et al.), or to decision bias and information merging from different sources (Connine, 1987; Norris et al., 2000; Samuel, 1981).

One of the factors that researchers have examined in attempting to explain both lexical and sentential context effects has been the timing of listeners' responses. If decision responses must be made as quickly as possible, both lexical and sentential context effects are significantly reduced (Miller & Dexter, 1988; Miller et al., 1984), suggesting that how much information is used for a decision may be under listener control. This decision-bias account contrasts with claims for automatic top-down feedback. In the absence of specific instructions about response speed, however, there is still variation in response speed (e.g., faster for easier decisions, slower for harder decisions). Lexical effects on word-initial phoneme decisions (such as *dice–tice*) tend to be larger in slower than in faster responses (Fox, 1984), because at the beginning of a word the fastest responses will be made before it is at all clear what the word is or could be. In contrast, lexical effects on word-final decisions (such as *fish–fiss, kiss–kish*) tend to be larger in faster than in slower responses (McQueen, 1991), because at the end of a word the lexical information is such as to predict the final phoneme very efficiently, and allow fast responding on the basis of lexical information alone. The word-initial timing difference is accounted for by both top-down and bottom-up accounts (for the relevant simulations with TRACE, see McClelland & Elman, 1986, pp. 27–28). This is not the case for the word-final difference, however. As McQueen pointed out, the Race model (Cutler et al., 1987) would indeed predict the observed pattern of larger lexical effects in fast responses (as would Merge and other autonomous models). In contrast, simulations with TRACE by McClelland (1987, pp. 12–13) predicted lexical effects to increase over time. It is important to understand that the difference between these predictions hinges to a large part on the availability of an unchanged record of the phonetic processing. TRACE over-rides this record with information flowing top-down, the effects of which accrue over time. This is why TRACE cannot predict the effects to be smaller in later responses. In autonomous models, the record of the phonetic processing is not overwritten, and an early bias towards a particular lexical choice can be modulated by phonetic evidence.

McQueen's (1991) argument was extended to sentential context by Van Alphen and McQueen (2001), using a manipulation of function words in Dutch modeled on an earlier experiment by Isenberg, Walker, and Ryder (1980) in English. Isenberg et al. constructed a continuum from a reduced article *the* to a reduced verbal marker *to* and found that listeners' judgments of the tokens differed when they appeared before *gold* versus *go*. The *the–to* syntactic contrast translates as *de–te* in Dutch, and Dutch has the added advantage that the infinitive verbal inflection is the same as a plural noun inflection, so that the following word form can be kept constant: *de schaatsen* means 'the skates', *te schaatsen* means 'to skate'. Van Alphen and McQueen found that categorization responses were indeed influenced by biasing context either preceding the sequence *the/to skate[s]* (*we are selling* versus *we dare*) or after it (*of my brother* versus *on speed skates*). Here the bias exercised by the preceding or following context is syntactic; *sell* favors a following noun phrase while *dare* favors a verb phrase, and the prepositional phrases likewise bias towards a noun and a verb, respectively.

Van Alphen and McQueen (2001) investigated the time course of the effect of this syntactic bias, by dividing responses into fast, medium, and slow reaction time (RT) ranges. This revealed the sentential effects to be weakest in the slow RTs, particularly with preceding context, and Van Alphen and McQueen again argued for a decision-bias account of their result.

Studies of ambiguous phonemes as created for these studies have greatly extended our understanding of information flow through the speech recognition system. Nevertheless, multiple stages in strictly regular progression along such a phoneme-to-phoneme continuum are more in service of the scientific need to determine how acoustic information is used than of a desire to mimic real speech under laboratory conditions. That is, in real life listeners are unlikely to be confronted with variation of this sort. They are frequently confronted, however, by ambiguities resulting from segment deletions and reductions in spontaneous speech. *What are you doing?* may be pronounced in such a way that it starts like the word *watch*, for example. Here too, reference to sentential context may be the only way to identify the intended meaning.

In the present study we investigate the contextual resolution of ambiguities arising from such natural reduction in spontaneous speech. The casual speech phenomenon that we use is word-final /t/-reduction (e.g., *post box* realized as [pəʊsbɒks]). This effect is common across languages and has been extensively studied in both English (Guy, 1980; Sumner & Samuel, 2005) and Dutch (Janse, Nooteboom, & Quené, 2007; Mitterer & Ernestus, 2006). Importantly, Mitterer and Ernestus found that this process is gradient, in that it is not always complete and allows varying degrees of reduction. This makes it particularly interesting as a test case for the present purposes, since it amounts to a natural continuum (from completely present /t/ to completely absent /t/) with which listeners will actually have had prior experience in their everyday listening.

Moreover, /t/-reduction also lends itself particularly well to the instantiation of a syntactic manipulation, since one of the uses of /t/ in Dutch is as a syntactic inflection: *zij rent* 'she runs' but *ik ren* 'I run'. The presence versus absence of a /t/ at the end of a tensed verb thus constitutes the difference between the third- and first-person singular present form. In some languages with /t/-reduction (e.g., German: Kohler, 1990), reduction of morphological affixes is discouraged, but this is not so in Dutch; morphological and non-morphological final /t/ are equally susceptible to reduction (see the production study in Mitterer & Tuinman, 2012). Thus for a task in which listeners must decide whether a word-final /t/ is present, verbs with a preceding pronoun provide an easily realizable syntactic constraint.

This was the structure of the present study. We presented listeners with a phonetic continuum from completely present to effectively absent /t/, using tokens for which listening data in the absence of constraining syntactic context had already been gathered by Mitterer and Ernestus (2006). These tokens appeared word-finally in verbs of which the stem ended either in /n/ or /s/. In Mitterer and Ernestus' study, and in subsequent studies by Mitterer and McQueen (2009) and Mitterer and Tuinman (2012), participants proved sensitive to the fact that natural reductions are more likely after /s/ than after /n/; this manipulation thus allowed a check that the present participants were attending to the phonetic context. Syntactic context took the form of either a first-person singular subject, inducing a /t/-absent bias, or a third-person singular subject, inducing a /t/-present bias. Listeners' task was to decide whether /t/ was present or absent. Following Van Alphen and McQueen (2001), across experiments we investigated the time course of syntactic context effects in different RT ranges, and we also varied whether the syntactic context preceded or followed the critical target-bearing word. In Experiment 1, the bias-generating sentence subject appeared before the critical verb.

2 Experiment 1

2.1 Method

2.1.1 Participants. Twenty-one members of the Max Planck Institute's subject pool, age range 18–33 and mean age 21.4, participated voluntarily in return for a small payment. All participants were native speakers of Dutch and none reported any hearing impairment.

2.1.2 Materials. Six Dutch verbs, three with a stem ending with /n/ (e.g., *zoen* 'kiss') and three with a stem ending in /s/ (e.g., *blaas* 'blow')[1] were used as target verbs. The sets of three were matched for lemma log frequency (Baayen, Piepenbrock, & Gulikers, 1995). Sentences containing the verbs preceded by either first-person or third-person pronouns and followed by one of five adverbs were constructed, and a male native speaker of Dutch (the same speaker as in Mitterer and Ernestus, 2006) recorded all sentences several times each. Table 1 shows the combinations of the target verbs and preceding and following words.

Table 1. Sentence frame for the stimuli in Experiment 1.

Connection word	Subject	Target verb	Adverb
Maar /mar/		blaas /blas/ (1.62)	nauwelijks /'nɑuʋələks/
	zij /zɛi/	kreun /krøn/ (1.46)	langzaam /'lɑŋzam/
		bloos /blos/ (1.21) ..t	moeizaam /'mujzam/
		zoen /zun/ (1.26) …Ø	soms /'sɔms/
	ik /k/	ren /rɛn/ (1.95)	vaak /'vak/
		kus /kys/ (1.75)	

Note. English translations are *Maar* 'but', *ik* 'I', zij 'she', *blaas* 'blow', *kreun* 'groan', *bloos* 'blush', *zoen* 'kiss', *ren* 'run', *kus* 'kiss', *nauwelijks* 'hardly', *langzaam* 'slowly' *moeizaam* 'with difficulty', *soms* 'sometimes', *vaak* 'often'. Note that the word order with the adverb after the verb is canonical in Dutch and the English word order 'I hardly run' would in fact be ungrammatical in Dutch. The numbers in brackets are the lemma log frequencies for the words taken from CELEX.

The utterances were then used as templates for synthesis via a Klatt synthesizer (Klatt, 1980). As in Mitterer and Ernestus (2006), formants and bandwidths were measured at the beginning, in the middle, and at the end of each segment and at major formant transition points within segments. When the formant could not be estimated reliably by Linear Predictive Coding (using Praat; Boersma & Weenink, 2005), values recommended by Klatt (1980) were chosen. In addition, nasalization for a nasal was carried over into the vowel before it was reduced completely, and anticipated for postvocalic nasals to match the natural utterances. Parameters for synthesis were generated by interpolating linearly between measurement points. Occasionally, values were altered in order to prevent clicks and other transients, which can occur if control parameters change too quickly. Amplitude values were iterated to imitate the amplitude envelope of the natural utterances. (The complete parameter set for all target verbs and the preceding and following contexts can be obtained from the authors.)

The verb stems were followed by one of five different synthesized signals for the coda (see Figure 1, and Mitterer & Ernestus, 2006). The first signal was similar to a full /t/ with a 25 ms closure and a 45 ms transient-frication sequence. To prevent an unnatural flat line in the signal, the closure was synthesized with 20 dB amplitude of frication (AF) and a 30 dB amplitude bypass (AB) of the parallel branch of the synthesizer. At the release of the initial burst, AB was set to zero, and AF increased to 40 dB, and decreased again to 15 dB at the end of the 65 ms signal. For this and all the other target signals, amplitude changes were loglinear in dB, to achieve a linear amplitude envelope. The transient-frication signal was dominated by the fifth and sixth formats (55 dB) starting at 5700 Hz (bandwidth of 500 Hz), and 7500 Hz (700 Hz), respectively. These formants fell to 5200 and 6850 Hz at 65 ms. The second, third, and fourth formant were held constant throughout the transient-frication sequence at 1434 Hz (200), 2212 Hz (500), and 3840 Hz (600), respectively, and their amplitude increased from 20 to 25 dB so as to mimic the increasing low-amplitude frication in the model [t].

The second coda signal was a 65 ms frication noise, as often found in /st/ codas. The AF started and ended at 15 dB with a 40 dB maximum at 30 ms. The settings for the second, third, and fourth formant were the same as the settings in the full [t]. The fifth and sixth formants started at 5130 Hz (bandwidth: 500 Hz) and 6750 Hz (700 Hz) and fell to 4620 and 6080 Hz at the end of the signal (As = 55 dB). The weak-frication signal, the third coda signal, was derived from this signal by reducing the overall amplitude of fricative noise by a factor of 5 (= 14 dB). The closure-only signal, the fourth coda signal, was synthesized with the same settings as for the closure of the full /t/.

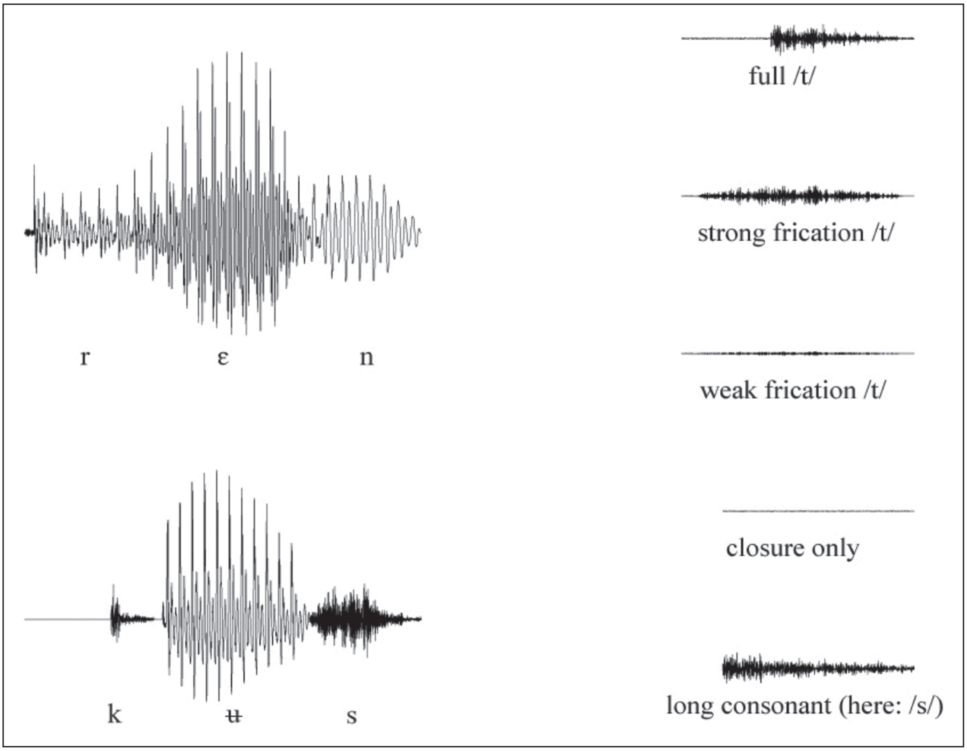

Figure 1. The different target signals presented in Experiment 1 and 2. See method sections (Section 2.1.2 for Experiment 1 and Section 3.1.2 for Exeriment 2) for more details.

The final coda signal was a coda that was not followed by any signal implying a /t/. Instead, it was created by elongating the consonant /n/ by 45 ms and /s/ by 65 ms. These durations were based on measured coda durations in the recording of the Dutch native speaker reading the test materials. These measurements also showed that consonants in consonant clusters are shorter than in simple codas.

Several adaptations were made in the interest of creating more natural-sounding utterances. The long-consonant target signals were adapted to the following adverb to make the transition as natural as possible between the target verb and the subsequent adverb. Thus, for the long-consonant /n/ the amplitude remained high if another nasal followed (e.g., if *nauwelijks* or *moeizaam* was the following adverb), but it fell before other phonemes. For the long-consonant /s/, the amplitude did not fall if the following sound was another /s/ (e.g., if *soms* was the following adverb), but decreased if the following sound was any other phoneme. A 25 ms closure was inserted after *Maar ik* and a 50 ms closure before verbs starting with a /k/ (*kus* and *kreun*) to make the synthesized materials more like the natural utterances. These alterations were all independent of the coda manipulation.

Overall, there were five coda signals for the target verbs (full /t/, strong frication /t/, weak frication /t/, closure only, long consonant), and all signals were combined with each of the six verbs, giving 30 distinct tokens for which participants had to decide whether they heard a coda with or without /t/. The 30 verb tokens were placed in 10 carrier sentences, created by combining the two possible sentence subjects (*Maar ik* … and *Maar zij*…) with five possible post-verb adverbs. This resulted in a total of 300 different stimulus sentences (see also Table 1).[2]

2.1.3 Procedure. The experiment was conducted on a standard PC running NESU experimental control software. Participants were tested one at a time in a normally lit soundproof booth. They wore Sennheiser closed headphones, sat at a comfortable reading distance from the computer screen, and had a two-button response box in front of them. Instructions in Dutch asked them to press the right-hand button if the sentence they heard contained the verb form shown in the upper-right corner of the screen, and to press the left-hand button if the sentence contained the verb form in the upper-left corner. They were instructed to ignore the fact that sentences could be ungrammatical with or without verb-final /t/, and to just respond on the basis of whether they had in fact heard /t/. After the instructions, four practice trials familiarized participants with the procedure.

The practice and experimental trials had the same structure. Each trial began with 150 ms of blank screen. The response alternative without coda /t/ (e.g., *blaas*) then appeared in the upper-left corner of the screen and the other alternative (e.g., *blaast*) appeared in the upper-right corner. After another 450 ms the sentence was played. A response, made by pressing one of the buttons, caused the other alternative to disappear from the screen, so that participants could see that their answer had been registered by the computer. If a participant did not react within 2.5 seconds from the onset of the target verb, a stopwatch appeared on the screen as a reminder to respond more rapidly. The post-response computer screen (with chosen response alternative or with stopwatch) stayed in place as feedback for one second before the start of the next trial. Participants heard all 300 stimulus sentences once, in a random order which differed for each participant. After every 50 trials participants could take a break and continue when ready.

The design thus included three independent variables: Coda Signal at the end of the target verb (full /t/, strong /t/ frication, weak /t/ frication, closure only, long consonant); Consonant preceding the target signal (stem-final /n/ or /s/); Syntax of the sentence (first-person pronoun making /t/ syntactically incorrect, third-person pronoun making /t/ syntactically correct). The single dependent variable was the percentage of /t/-responses in each cell of the design.

2.2 Results

2.2.1 Analysis of number of /t/-responses. RTs faster than 100 ms (after offset of the target verb) were discarded from this analysis. Prior to the analyses, the RTs were corrected to reflect the response time from offset of the target verb. All cases in which no response was given were also removed from the analyses. In total, 1.4 % of trials were removed.

The mean percentages of /t/-responses, shown in Figure 2, were subjected to a repeated-measures analysis of variance (ANOVA) with Syntax, Coda Signal, and Consonant as predictors. The data were log-odds transformed for this analysis, with zeros and ones replaced by 0.01 and 0.99, respectively. As can be seen in from Figure 2, listeners were more likely to report the presence of a /t/ if this interpretation led to a syntactically correct sentence [square symbols show higher values than round symbols; Syntax: $F(1, 20) = 49.77$, $p < 0.001$; *ik*: 65.3%, *zij*: 82.5%] and if the Coda Signal carried more acoustic information for the presence of a /t/ [higher values on the left of the figure than on the right; $F(4, 80) = 122.68$, $p < 0.001$]. Figure 2 also clearly shows that more /t/-responses were given after a preceding /s/ than after a preceding /n/ [dark symbols show higher values than light symbols; $F(1, 20) = 46.58$, $p < 0.001$; /s/: 82.1%, /n/: 65.8 %]. Interactions of Syntax and Consonant, $F(1, 20) = 26.68$, $p < 0.001$, and Syntax and Coda Signal, $F(4, 80) = 6.51$, $p < 0.001$, were significant; other interactions did not reach significance [Consonant by Coda Signal: $F < 1$; Syntax by Consonant by Coda Signal: $F(4, 80) = 1.46$, $p > 0.1$].

To explore the interaction between Syntax and Consonant, we examined the effect of Syntax separately given Consonant /n/ or /s/. The Syntax effect was significant in both cases, $p_{max} < 0.001$,

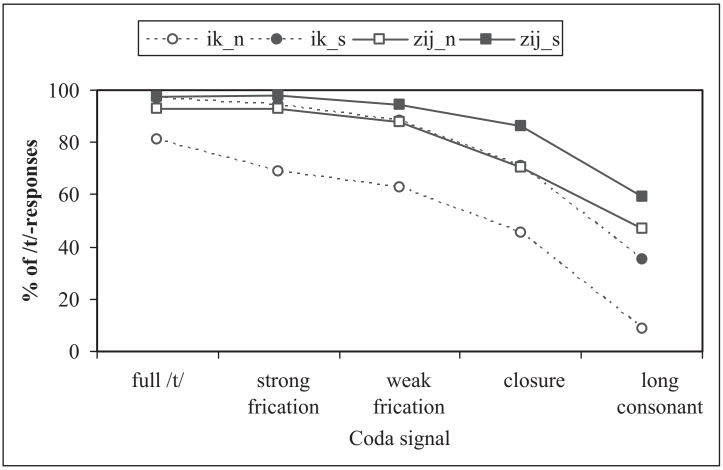

Figure 2. Percentages of /t/-responses in Experiment 1 for each Coda Signal. The lines indicate the two different Preceding Contexts /n/ and /s/ and the two different Syntax conditions, first-person singular *ik* and third-person singular *zij*.

but its size (the difference in percentage of /t/-responses given *ik* versus *zij*) was significantly larger after /n/ (24.6%) than after /s/ (9.8%).

Effects of Syntax as measured in *t*-tests were also significant for all levels of the /t/-Ø series, $p_{max} < 0.01$; the interaction of Syntax by Coda Signal thus again arose from differences in the size of the effect. A one-way ANOVA across all five Coda Signal levels revealed a main effect of Syntax, $F(1, 20) = 6.03$, $p < 0.001$, with post-hoc tests showing *ik–zij* differences to be smallest with the full /t/ signal and largest with a long consonant; the three intermediate signals (strong frication /t/, weak frication /t/, and closure) did not differ significantly from one another.

The mean percentage of /t/-responses for the long-consonant Coda Signal /s/ was 37.4% when a fricative ('vaak' or 'soms') followed, and only 22.5% with other following contexts. A possible explanation is that the amplitude variations in the resulting long stretch of frication led listeners to interpret it as a frication /t/, resulting in more /t/-responses. This would indeed mirror patterns of production, given that Mitterer and Ernestus (2006) found that deletion of word-final /t/ becomes more likely as the first segment of the following context becomes less sonorant.

2.2.2 RT range analyses. To track the influence of Syntax over time, a RT split was applied. Because RTs differed across conditions (from 526 ms for *zij*, /s/, and strong frication to 760 ms for *ik*, /n/, and strong frication) we did not use a fixed cut-off, but ranked RTs for each participant individually and for each of the 20 cells of the design as McQueen (1991) and Van Alphen and McQueen (2001) had done. From the ranked RTs we then selected the 40% fastest responses (Fast range) and the 40% slowest responses (Slow range), omitting RTs in the mid-range 40–60%. This difference from the two earlier studies (which each compared fast, medium, and slow ranges) was necessary because our more complex factorial design, with many more cells within which RTs were individually ranked for each participant, did not yield enough trials to allow a three-way split per cell. Therefore, we generated two well-separated bins by excluding the smaller 40–60% range and comparing the remaining 80% of the data as a larger fast against a larger slow bin. A repeated-measures analysis was performed on the percentages of /t/-responses with Syntax, Coda Signal, Consonant, and RT range as predictors.

This analysis showed that responses in the fast RT range were significantly more often a /t/-present response than responses in the slow range, RT range: $F(1, 20) = 16.12$, $p = .001$; fast: 76.1%; slow: 71.2%; see Figure 3. Syntax also interacted significantly with RT range, $F(1, 20) = 10.41$, $p < .01$, and with Consonant by Coda Signal by RT range, $F(4, 80) = 3.36$, $p < 0.05$; all other interactions with the RT range were insignificant ($p > .05$). The effect of Syntax was significant for fast and slow RT ranges separately, $p_{max} < 0.001$, so that the interaction of Syntax and RT range again reflects effect size differences, here a significantly larger Syntax effect for the fast RTs (21.2%) than for the slow RTs (13.4%), $t(20) = 3.22$, $p < 0.01$. Thus the sentential context effect, syntax in this case, decreases over time.

The four-way interaction of Syntax by RT range by Consonant by Coda Signal can be understood from Table 2, in which it can be seen that the effect of Syntax is larger in fast RTs than in slow RTs for nine of the ten combinations of Consonant and Coda Signal, but is larger in slow RTs for the remaining one ('closure only' with /s/).

2.3 Discussion. Three clear results emerged from this experiment. Firstly, the probability of /t/-responses was influenced by the sentence syntax: more /t/-responses were given to sentences that were /t/-biased by the third-person pronoun *zij* than to sentences that were biased against /t/ by the first-person pronoun *ik*. Secondly, the two findings of Mitterer and Ernestus (2006) were replicated: listeners took the preceding consonant into account, giving more /t/ responses after /s/ than after /n/, and they took the subphonemic detail of coda signals into account in that they gave more /t/-responses when acoustic cues supported the presence of a /t/, and fewer /t/-responses when acoustic cues did not support /t/. Thirdly, the influence of syntactic context on the identification of /t/ was weaker in slower than in faster responses.

In Experiment 2, following the example of Van Alphen and McQueen (2001), we tested whether the time course of the syntactic context effect would change if the sentence word order is such that the syntactic information is presented *after* the target sound.

3 Experiment 2

In this experiment, we changed the word order so that the syntactic bias occurred after the critical target word in the sentence frame. Because of the constraints of Dutch grammar, the crucial

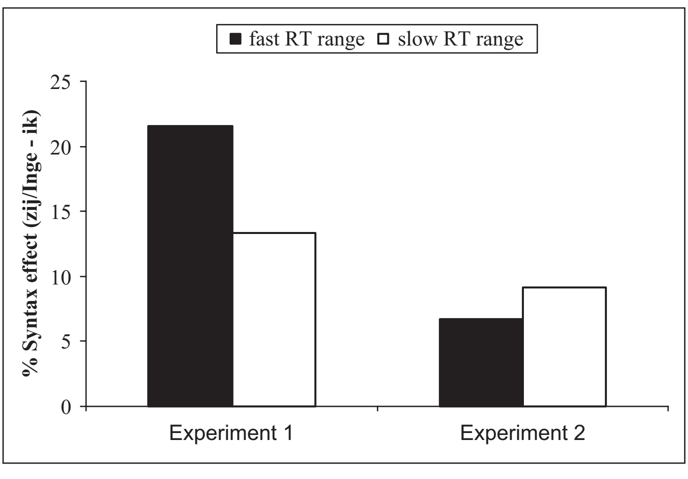

Figure 3. Syntax effect in fast and slow reaction time (RT) ranges in Experiment 1 and Experiment 2.

Table 2. Effect of Syntax on the mean percentage of /t/-responses for different Preceding Contexts /n/ and /s/, Coda Signals in the fast and slow reaction time (RT) ranges in Experiment 1.

Preceding Context	Coda Signal	Fast RT range Syntax effect (zij-ik)	Slow RT range Syntax effect (zij-ik)	Syntax × RT range (*F*-values and significance)
	Full t	13.5	7.9	1.16
	Strong frication	29.5	18.2	6.46*
n	Weak frication	29.4	27.0	0.12
	Closure	35.1	13.5	6.17*
	Long consonant	43.9	29.4	4.85*
	Full t	1.7	0	0.17
	Strong frication	4.3	0.8	0.66
s	Weak frication	7.3	7.0	0.43
	Closure	5.9	20.2	2.44
	Long consonant	41.5	10.3	8.73 *

syntactic information (e.g., the subject of the sentence in the form of a pronoun or common name) cannot appear very late in the sentence. In Dutch, the subject and the finite verb are usually directly adjacent (unless a subordinate clause intervenes). However, they can trade places. In sentences starting with an adverb, the order of the subject and verb is obligatorily inverted: *Soms ren(t) ik naar huis* (literally: Sometimes run(s) I home = 'Sometimes I run home'). The pronoun *ik* cannot be placed further from the verb *ren*. To keep the phonological context following the reduced /t/ identical, the common name *Inge* was therefore used instead of the pronoun *zij* ('she'), so that the parallel sentence with a /t/-bias is *Soms ren(t) Inge naar huis* (literally: Sometimes run(s) Inge home = 'Sometimes Inge runs home').

With this manipulation, we can examine the influence of syntactic information that is presented later and can thus only be processed later. We can also again examine whether syntactic information can only influence phonetic decisions in a limited time window, and whether syntactic context effects rise, decrease or remain stable over time.

3.1 Method

3.1.1 Participants. Twenty-one members of the Max Planck Institute's subject pool, age range 19–25 and mean age 21.1, participated in return for a small payment. All were native speakers of Dutch, none reported any (history of) hearing problems and none had taken part in Experiment 1.

3.1.2 Materials, procedure, and design. The same target verbs were used as in Experiment 1. Again the verbs were placed in sentences, with in this case the subject of the sentence after the verb. The same speaker as in Experiment 1 recorded the sentences, schematized in Table 3, several times. As in Experiment 1, the recorded utterances were used as templates for synthesis. For the words that also occurred in Experiment 1, the parameter settings used in Experiment 1 were used as a starting point and adjusted where needed to model the natural utterances recorded for Experiment 2. The main difference for the target verbs and adverbs across the two experiments was in duration and pitch contour, resulting from the syntactic changes; for example, the adverbs were now

Table 3. Sentence frame for the stimuli in Experiment 2.

Adverb	Target verb	Subject	Place/object
Nauwelijks /'nɑuʋələks/	blaas /blas/ (1.62)	ik /ɪk/	een noot / en not/
Langzaam /'lɑŋzam/	kreun /krøn/ (1.46)		op maandag /ɔp mandɑx/
Moeizaam /'mujzam/	bloos /blos/ (1.21) ..t		op school /ɔp sxol/
soms /'sɔms/	zoen /zun/ (1.26) …Ø	Inge /ɪŋə/	mijn moeder /mɛin mudər/
vaak /'vak/	ren /rɛn/ (1.95)		naar huis /nar hœys/
	kus /kys/ (1.75)		haar tante /har tɑntə/

Note. English translations are *nauwelijks* 'hardly', *langzaam* 'slowly' *moeizaam* 'with difficulty', *soms* 'sometimes', *vaak* 'often', *blaas* 'blow', *kreun* 'groan', *bloos* 'blush', *zoen* 'kiss', *ren* 'run', *kus* 'kiss', *ik* 'I', *Inge* 'Inge' (common Dutch first name), *een noot* 'a note', *op maandag* 'on Monday', *op school* 'at school', *mijn moeder* 'my mother', *naar huis* 'home', *haar tante* 'her aunt'. The numbers in brackets are the lemma log frequencies for the words taken from CELEX.

utterance-initial and thus shorter than in Experiment 1, where they had been utterance-final and subject to phrase-final lengthening. Otherwise, stimuli were created as in Experiment 1 and the same five coda signals were used. Procedure and design were as for Experiment 1.[2]

3.2 Results

3.2.1 Analysis of the number of /t/-responses. As in Experiment 1, we discarded RTs faster than 100 ms (after offset of the target verb), as well as all cases in which no response was given, leading to removal of 1.6 % of the trials.

The mean percentages of /t/-responses are shown in Figure 4. A repeated-measures ANOVA on the log-odds transformed percentages of /t/-responses was again performed with Syntax, Coda Signal, and Preceding Context as predictors. As Figure 4 clearly indicates, the consonant /s/ elicited more /t/-responses than the consonant /n/, $F(1, 20) = 65.61$, $p < 0.001$; /s/: 71.8%, /n/: 35.8%. Listeners were also more likely to report the presence of a /t/ if this interpretation led to a syntactically correct sentence, Syntax: $F(1, 20) = 25.06$, $p < 0.001$; ik: 49.9%, Inge: 57.7%, and if the Coda Signal carried more acoustic information for the presence of a /t/, $F(4, 80) = 85.62$, $p < 0.001$.

In addition, there were significant interactions of Syntax and Consonant, $F(1, 20) = 17.97$, $p < 0.001$. This was due to the fact that the effect of Syntax was significant for /n/, $t(20) = 6.6$, $p < 0.001$, but not for /s/, $t(20) = 1.3$, $p > 0.1$. There was also an interaction of Coda Signal and Consonant, $F(4, 80) = 23.24$, $p < 0.001$. This interaction was again due to a difference in size of the effect over the continuum: The effect of Consonant was significant for all levels of the factor Coda Signal, $t_{min}(20) = 2.3$, $p < 0.05$. Post-hoc comparisons on the effect sizes of Consonant on all levels of the Coda Signal factor showed significantly larger effect sizes in the middle of the continuum than at the margins. All other interactions failed to reach significance.

3.2.2 RT range analyses. Again we examined the syntax effect over time with a RT split as in Experiment 1. A repeated-measures analysis on the /t/-response percentages with Syntax, Coda Signal, Consonant, and RT range as predictors revealed neither a main effect of RT range, $p > 0.9$, nor any significant interactions of RT range with Syntax, $p > 0.3$. While RT range interacted significantly with Consonant, $F(1, 20) = 14.85$, $p < 0.005$, with Coda Signal, $F(4, 80) = 7.60$, $p < 0.001$, and with Consonant by Coda Signal, $F(4, 80) = 2.82$, $p < 0.05$, the effect of Syntax was not significantly different over time (see Figure 3). Given that the effect of Syntax was significant with preceding /n/ but not /s/ (although there was a difference in the same direction in the /s/-context

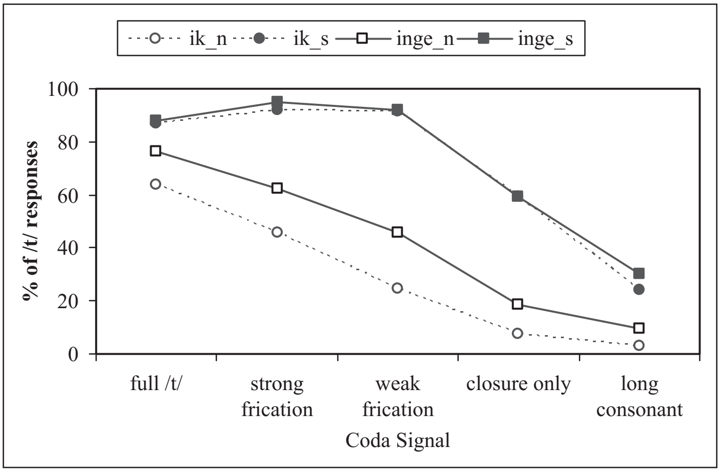

Figure 4. Percentages of /t/-responses in Experiment 2 for each Coda Signal. The lines indicate the two different Preceding Contexts /n/ and /s/ and the two different Syntax conditions, first-person singular *ik* and third-person singular common name *Inge*.

as well), we repeated the RT range analysis on the data for the /n/-context only. This did not lead to different conclusions; RT range interacted significantly with Coda Signal, $F(4, 80) = 4.70$, $p < 0.01$, but there was neither a main effect of RT range, $F(1, 20) = 3.35$, $p = 0.082$, nor a significant interaction of RT range with Syntax, RT range × Syntax: $F(1, 20) = 2.66$, $p > 0.1$; RT range × Syntax × Coda Signal: $F(1, 20) = 1.00$, $p > 0.1$.

3.3 Discussion

Once again, syntax, preceding consonant, and subphonemic detail influenced the response pattern. Even though the crucial syntactic information appeared after the ambiguous signal that had to be identified, more /t/-responses were given for sentences with the subject *Inge* than for sentences with the subject *ik*. Listeners also gave more /t/-responses after /s/ than after /n/, and more /t/-responses for coda signals with strong evidence for an underlying /t/ than for codas with weak evidence for an underlying /t/.

In contrast to Experiment 1, however, in this experiment the influence of Syntax did not differ for fast and slow responses. A repeated-measures analysis of both experiments together showed an interaction between Syntax, RT range, and Experiment, $F(1, 40) = 7.86$, $p < 0.01$; thus the patterning of Syntax with the RT ranges differed significantly in the two experiments. The Experiment variable in fact interacted with every other independent variable. A larger Syntax effect in Experiment 1 (17.3%) than in Experiment 2 (7.9%) produced an Experiment–Syntax interaction, $F(1, 40) = 10.87$, $p < 0.01$. In Experiment 1, 5.0% more /t/-responses fell in the fast range than in the slow range, while the difference between the two RT ranges in Experiment 2 was 0.1%; this led to an interaction of Experiment with RT range, $F(1, 40) = 6.74$, $p < 0.05$. The interactions of Experiment with Coda Signal, $F(4,160) = 3.50$, $p < 0.05$, and Consonant, $F(1, 40) = 11.0$, $p < 0.005$, both stem from the cross-experiment difference in phonetic context following the target site. The word following the crucial ambiguity in the present Experiment 1 was always an adverb beginning with a consonant, but the following word in Experiment 2 was either *ik* or *Inge*, beginning with a vowel. Mitterer and Ernestus (2006) found that /t/-reduction is more likely before consonants than before vowels, and Mitterer and McQueen (2009) showed that listeners are sensitive to this likelihood. A general bias towards /t/-responses appeared in Experiment 1 (Figure 2) but less so in Experiment 2 (Figure 4), that is, listeners again compensated for the likelihood of a following context triggering /t/-reduction. Recall also that the effect of preceding consonant (/n/ versus /s/) was larger in Experiment 2 (36%: 71.8% /t/-responses after /s/, 35.8 % after /n/) than in Experiment 1 (16%: 82.1% after /s/, 65.8 % after /n/). This suggests that stimuli in the prevocalic position are more ambiguous, allowing a preceding consonant to capture more of the response variance than had been taken up by the effect of the same consonant on the (preconsonantal) judgments made in Experiment 1.

4 General discussion

The results of our two experiments motivate four conclusions. Firstly, a continuum of realistic phonological reduction elicits a graded response pattern from listeners, just as has been observed in experiments with artificially constructed phoneme-to-phoneme continua. Secondly, presence versus absence of a syntactic inflection can be judged in the same way as phonemic category membership. Thirdly, the response continuum that results from the listeners' judgments of presence versus absence along a continuum of reduction is subject to syntactic context effects analogous to those that have been observed for lexical identity biased by sentence context. Fourthly, the timing of the use of this syntactic context for making these judgments about word-final inflection differs as a function of whether the contextual information precedes or follows the site of the inflection, and patterns in a similar way to syntactic context effects on word judgments with artificially manipulated phoneme continua. In short, all four of these results are new, but all four also confirm the conclusions that have arisen from prior research with artificial phoneme continua. Our results thus leave intact the over-arching conclusion that the pattern of results across all these studies is consistent with decision-bias models, but not with models in which top-down lexical or syntactic influences on lower-level processes are mandatory.

The manipulation of phoneme identity from one category endpoint to another can present listeners with stimuli unlike anything they have ever heard before, particularly if place or manner of articulation is manipulated (it is arguable that voicing manipulations may succeed in mimicking possible realizations in particular contexts). Under such circumstances, listeners making a decision about a truly ambiguous token have little option other than to base their decision on whatever information is provided by the sentence context (lexical, syntactic, and pragmatic). The speech manipulation underlying our stimuli, however, was one with which listeners have acquired natural experience. The tokens were directly modeled on the perceptual study of Mitterer and Ernestus (2006), and they in turn devised those tokens on the basis of corpus research using the extensive Corpus of Spoken Dutch (Oostdijk, 2000), in which realizations along the chosen continuum were found to occur both within and across speakers. When listeners are confronted, in natural conversation, with a stretch of speech that is truly ambiguous as to whether or not a /t/ is present, will they too resort to the sentential context to support their decision? Our results suggest that they will.

We achieved true ambiguity in our study by using synthetic stimuli. While we might in principle rather have created stimuli from naturally occurring tokens, and corpora reveal that across time and across speakers, all the requisite tokens along a continuum do occur, in practice individual speakers cannot produce reduced tokens exactly placed along a continuum 'on cue' as necessary for our experiment. Thus synthetic speech allows us to generate stimuli that mimic the properties of spontaneous speech quite closely, while still controlling all the required source properties. Note here that the effects of phonological context and phonetic detail reported by Mitterer and Ernestus (2006) with synthetic materials of the type also used here have been replicated with (edited) natural speech (Mitterer & McQueen, 2009). Although synthetic speech has an unusual voice quality, this does not seem to interfere with how listeners deal with phonetic ambiguities in it.

In by far the majority of ambiguity-resolution experiments within the tradition to which our study contributes, listeners' judgments have concerned word identity (e.g., *bath* versus *path, goat* versus *coat, to* versus *the*, or a real-word versus a non-word endpoint: Borsky et al., 1998; Ganong, 1980; Isenberg et al., 1980; McQueen, 1991; Rohde & Ettlinger, 2012; Van Alphen & McQueen, 2001). Our listeners made a judgment that concerned the presence versus absence of an inflection. Again, such morphemic uncertainties are likely to have been encountered by our listeners in real life, not only because reduction of morphological /t/ in Dutch is reasonably common (Mitterer & Ernestus, 2006), but also because morphemic errors, including affix deletion, occur quite frequently as slips of the tongue (Garnham, Shillcock, Brown, Mill, & Cutler, 1981). In both our experiments, more /t/-responses were made in sentences with third-person subjects (consistent with a t-inflection) than in sentences with first-person subjects (inconsistent with t-inflection). Thus our results suggest that in cases where the presence of an inflection is unclear, relevant syntactic context will be exploited to clear up listeners' momentary uncertainty.

In a relatively realistic confrontation with speech ambiguity, we thus see that the syntactic context is used rapidly to inform decisions that in essence concern the presence of phonetic evidence. The crucial question is then how the influence of such context on phonetic decisions comes about: by automatic top-down flow of information from higher levels of processing to lower, or by integration of information from different levels to inform decision-making? As laid out in the introduction, one way to address this question is to take into account the timing of the decisions, that is, comparing fast and slow decisions. A model such as TRACE (McClelland & Elman, 1986) predicts that slow decisions would be subject to more top-down influence than fast decisions, because, in TRACE, lexical influences build up, over-ride uncertain evidence from below, and become stronger with time. Assuming a decision-bias model, McQueen (1991) argued that the accumulated lexical information alone may support phonemic decisions when an ambiguous sound appears in the word-final position. A decision-bias model also allows, in the case of word-final ambiguity, for slower/later decisions to be based on integration of information from the lexicon and from the signal, and for the contribution of information from the lexical context to be, in consequence, proportionally smaller. A decision-bias model further allows influence of any kind of context to decrease as well as increase. Consider that in Merge (Norris et al., 2000), phoneme decisions can be biased by syntactic context, but this bias can be undone if not supported by incoming pre-lexical information. The strength of the decision bias depends inter alia on when the contextual information is available to be merged with information from other sources. Importantly, as in any autonomous model, the pre-lexical representations in Merge are never altered by such bias.

Again, our results contribute further useful evidence. We created a situation where in each of our experiments the contextual information was not lexical but syntactic, and the ambiguous sound was word-final, but either decision would create an existing form of the presented verb. Syntactic information suggesting which verb form would be correct was supplied in Experiment 1 by a pronoun that preceded the verb, making the situation comparable in availability of information to the word-final lexical manipulation in, for example, McQueen's *kiss–kish* or *fiss–fish*. (It is somewhat less comparable to Van Alphen and McQueen's (2001) closed-class ambiguity, since disambiguation between *te* and *de* was not fully forced by the preceding semantic context they used.) In Experiment 1, in which /t/-judgments were influenced by the person of a preceding pronoun, fast identification decisions showed more influence of that syntactic information than slow decisions: the difference due to preceding pronoun was significant in both fast and slow RT ranges, but it was significantly larger in the responses falling into the listeners' range of fast RTs than in those falling into their range of slow RTs. This replicates the results obtained by McQueen and again points towards a decision-bias account of contextual influence, in this case syntactic.

What then is the situation in our Experiment 2, in which syntactic context in the form of the sentence subject appeared after the site of the phonetic ambiguity? Again our materials differ from those of Van Alphen and McQueen (2001), where several words separated the ambiguous sound from its disambiguation. In our materials, the disambiguations appeared after just the one constant initial vowel of *Inge* and *ik*, so that they more resemble word-initial lexical manipulations such as *butt–dutt* (Fox, 1984) or *bath–path* (Miller et al., 1984); there, too, the disambiguation arrived within a syllable. Effects of context were larger in slower than in faster responses in these two studies, both when the context was purely lexical, as in Fox's *butt–dutt*, or when it was syntactic, as in Miller et al.'s *bath–path* preceded by *hot water for the–* versus *jog along the–*.

Despite resembling those earlier studies on the information availability dimension, though, Experiment 2 did not produce exactly the pattern of responses previously found. The syntactic context effect was, first of all, much weaker than in Experiment 1, as the cross-experiment analysis showed; also, it appeared with verbs of which the stem ended in /n/, such as *ren* and *zoen*, but essentially did not appear with the /s/-final verbs such as *blaas* and *kus*. As suggested in discussion of Experiment 2, the following phonological context is the likely source of this latter pattern; in real life /t/-reduction occurs more often before consonants than before vowels, and listeners are sensitive to this, so the post-target consonants in Experiment 1 provided more opportunity for listeners to vary their expectations in accord with the syntax than did the post-target vowels in Experiment 2.

The residual effect of the syntactic context in Experiment 2 then did not pattern differently in fast versus slow responses. As noted in the introduction, larger context effects in slower responses, as found both by Fox (1984) for the lexical context case and Miller et al. (1984) for the syntactic context case, are predicted by a model such as TRACE as well as by a decision-bias model in which fast responses are based solely on information from the signal. A decision-bias model also copes with variable patterning of contextual effects across different experiments as a function of strategic adaptations by participants to the experimental task demands, whereas such variability is not predicted by models in which top-down context effects are automatic. Note that this kind of variability across experiments that differ in when contextual information is available appeared not only in our study, but also in Van Alphen and McQueen's (2001) study with artificial phoneme continua in closed-class words. Just as here, they found that effects of context preceding the ambiguity became weaker over time, but effects of context following the ambiguity did not change significantly over time; a decision-bias model was their choice for explaining this asymmetry. Further evidence comes from a recent study by Rohde and Ettlinger (2012) in which listeners' judgments of whether an ambiguous pronoun was *he* or *she* were affected by the pragmatics induced by a preceding verb. Instead of collecting RTs, Rohde and Ettlinger investigated the timing patterns in their study with a separate gating experiment in which listeners chose between *he* and *she* given the preceding sentence context plus the crucial pronoun in five steps of increasing size. Pragmatic effects appeared in early gates where the pronoun was not actually audible, but no such effects were observed on responses to the gate at which the onset consonant plus its transition into the vowel were first fully available. This shows that the strong pragmatic bias that clearly affected responses in this gating task did not cause the information in the ambiguous signal to be overwritten. Pragmatic effects then reappeared at a later gate (at which the ambiguous word could be integrated in full with the preceding context). This variable pattern is again clearly in line with our decision-bias account.

Context effects on our interpretation of perceptual input in the world are real, pervasive, and multifarious (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006; Mitterer, Horschig, Müsseler, &

Majid, 2009; Schutz & Lipscomb, 2007). Speech processing too shows many instances of this: listeners use semantic context to anticipate upcoming words (see, e.g., Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005), they use prosodic context to interpret syntactic phrase structure (see, e.g., Schafer, Carter, Clifton, & Frazier, 1996), and, as the literature we have surveyed abundantly shows, they use lexical and all varieties of sentential context to resolve phonetic ambiguity. The pattern of results across experiments in this latter area so far is better accounted for by decision-bias models than by models in which top-down effects over-ride bottom-up input; we propose that the same is true of our results. We have shown that listeners deal with ambiguity continua of the kind confronting them in everyday life by making use of whatever disambiguating information is available at the time. Our study has thus confirmed that ambiguity studies using artificial phoneme-to-phoneme continua can lay claim to ecological validity, and has further emphasized that the way in which decisions about speech input are made is under listener control to the extent that it can vary with the nature of an experimental task.

Acknowledgements

We thank James McQueen, Mirjam Ernestus, Larissa Ranbom, and an anonymous reviewer for helpful comments. The first author is now at the Centraal Bureau voor de Statistiek, Den Haag. The second author is now at the University of Malta, Faculty of Media and Knowledge Sciences, Department of Cognitive Science, Msida, Malta.

Funding

This work was supported by a Max Planck Society PhD scholarship to the first author.

Notes

1. Audience response to these examples compels us to remind the readers that this is Dutch, not English. As in many European countries, a kiss is a normal social greeting in The Netherlands, and the metaphors of any one language do not translate to another. In other words, our materials involved neutral and unremarkable sentences. Likewise, the syntax of Dutch differs from that of English (see also the caption to Table 1). In the present text, English glosses in single quotes are translations; literal transliterations have no quotes.
2. The full audio materials can be downloaded at http://www.mpi.nl/people/cutler-anne/research and http://www.holgermitterer.eu/research.html#stimuli.

References

Baayen, H. R., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (release 2): Linguistic Data Consortium.

Boersma, P., & Weenink, D. (2005). Praat 4.2. Amsterdam, The Netherlands: University of Amsterdam.

Borsky, S., Shapiro, L. P., & Tuller, B. (2000). The temporal unfolding of local acoustic information and sentence context. *Journal of Psycholinguistic Research*, *29*, 155–168.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat": The effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustical Society of America*, *103*, 2670–2676.

Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition - The role of sentence context. *Journal of Memory and Language*, *26*, 527–538.

Connine, C. M. (1990). Recognition of spoken words with ambiguous word initial phonemes. *Bulletin of the Psychonomic Society*, *28*, 497–497.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition - Temporal and linguistic constraints. *Journal of Memory and Language*, *30*, 234–250.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*, 141–177.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 526–540.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110–125.

Garnham, A., Shillcock, R. C., Brown, G. D. A., Mill, A. I. D., & Cutler, A. (1981). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics*, *19*, 805–817.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes*, *2*, 613–656.

Guy, G. R. (1980). Variation in the group and the individual. In W. Labov (Ed.), *Locating language in time and space* (pp. 1–36). New York, NY: Academic Press.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*, 1367–1368.

Isenberg, D., Walker, E. C. T., & Ryder, J. M. (1980). A top-down effect on the identification of function words *Journal of the Acoustical Society of America*, *68*, 48.

Janse, E., Nooteboom, S., & Quené, H. (2007). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes*, *22*, 161–200.

Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971–995.

Kohler, K. J. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. H. A. Marchal (Ed.), *Speech Production and Speech Modelling* (pp. 69–92). Dordrecht, The Netherlands: Kluwer.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention & performance XII: The psychology of reading* (pp. 1–36). London, UK: Erlbaum.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization - stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 433–443.

Miller, J. L., & Dexter, E. R. (1988). Effect of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 369–378.

Miller, J. L., Green, K., & Schermer, T. M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception & Psychophysics*, *36*, 329–337.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers lenite: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, *34*, 73–103.

Mitterer, H., Horschig, J. M., Müsseler, J., & Majid, A. (2009). The influence of memory on perception: It's not what things look like, it's what you call them. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*, 1557–1562.

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*, 244–263.

Mitterer, H., & Tuinman, A. (2012). The role of native-language knowledge in the perception of casual speech in a second language. [Original Research]. *Frontiers in Psychology*, 3, 249.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*, 357–395.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–324.

Oostdijk, N. (2000). The Spoken Dutch Corpus, overview and first evaluation. In *Proceedings of LREC-2000, Athens* (*Vol. 2*, pp. 887–894).

Rohde, H., & Ettlinger, M. (2012). Integration of pragmatic and phonetic cues in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 967–983.

Samuel, A. G. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1124–1131.

Schafer, A., Carter, J., Clifton, C., & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes*, *11*, 135–163.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration. *Perception*, *36*, 888–897.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, *52*, 322–338.

Van Alphen, P., & McQueen, J. M. (2001). The time-limited influence of sentential context on function word identification. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 1057–1071.

Van Berkum, J. J. A., Brown, C. M., 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*, 443–467.