The Mental Lexicon Is Fully Specified: Evidence From Eye-Tracking

Holger Mitterer

Max Planck Institute for Psycholinguistics

Four visual-world experiments, in which listeners heard spoken words and saw printed words, compared an optimal-perception account with the theory of phonological underspecification. This theory argues that default phonological features are not specified in the mental lexicon, leading to asymmetric lexical matching: Mismatching input ("pin") activates lexical entries with underspecified coronal stops ('*tin*'), but lexical entries with specified labial stops ('*pin*') are not activated by mismatching input ("tin"). The eye-tracking data failed to show such a pattern. Although words that were phonologically similar to the spoken target attracted more looks than unrelated distractors, this effect was symmetric in Experiment 1 with minimal pairs ("tin"-"pin") and in Experiments 2 and 3 with words with an onset overlap ("peacock" - "teacake"). Experiment 4 revealed that /t/-initial words were looked at more frequently if the spoken input mismatched only in terms of place than if it mismatched in place and voice, contrary to the assumption that /t/ is unspecified for place and voice. These results show that speech perception uses signal-driven information to the fullest, as predicted by an optimal perception account.

Introduction

The listener has to solve at least two overwhelming problems. First, he or she needs to store a huge number of words (~5 * 10^4) in such a way that they are all easily accessible during comprehension and production. Second, in comprehension, the listener needs to recognize words even if the input deviates from the canonical pronunciation. In contrast to written language, such deviations occur regularly in spontaneous dialogues (Johnson, 2004).

The theory of a featurally underspecified lexicon (FUL) (Eulitz & Lahiri, 2004; Lahiri & Marslen-Wilson, 1991; Lahiri & Reetz, 2002) provides—prima facia—good answers to both of these problems. According to this theory, the basic unit of lexical entries is the phonological feature, so that, for example, the representation of the last segment of the word *rum* is specified as [Nasal] and [Labial]. Importantly, predictable default features are not specified. As a consequence, lexical representations of words are underspecified, that is, not all features are specified.[1] The last segment of

The author thanks Robbert van Sluis for helping to prepare and for running these experiments. I also thank James McQueen and Janise Farrel Shanley for comments on an earlier version of the manuscript.

Correspondence concerning this article should be addressed to Holger Mitterer, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands. Electronic mail may be sent to Holger(dot)Mitterer(at)mpi(dot)nl.

the word *run* is therefore only specified as [Nasal], because coronal place of articulation of the /n/ is assumed to be default and is as such not specified.

Because the left-out, unspecified features are not stored in memory, underspecification reduces the memory load for the mental lexicon. Proponents of underspecification have often followed a linguistic tradition that assumes that the mental representations should be as parsimonious as possible and hence not contain any predictable features (Archangeli, 1988; Cairns, 1988; Scharinger, 2009). It is therefore better to leave frequent coronals unspecified than leaving less frequent labials (/b/,/p/,/f/, ...) unspecified, because alveolar segments are the more frequent than labial segments (Paradis & Prunet, 1991). Underspecification of coronals would therefore reduce the memory load substantially. However, more recent research indicates that the human memory capacity is rather large (see Goldinger, 1998, for an overview), so that the efficiency argument of underspecified representation is no longer strong.

The assumption of underspecification in the FUL model also leads to three types of matching relations between input and lexicon, which may help the listener to deal with variation in casual speech (Lahiri & Reetz, 2002). First, if a feature is specified in the lexicon and occurs in the input, there is a "match" (e.g., "flame" matches *flame*); second, if a feature is represented and another feature occurs in the input, there is a "mismatch" (e.g., "flane" mismatches *flame*); third and most critically, if a feature is not specified in the lexicon, there is "no mismatch", and this is independent of the input. That is, both "green" and "greem" match the lexical representation for *green*, with a "no mismatch" for the final /n/. The lexicon specifies only a nasal in the final position, and both inputs "green" and "greem" end in a nasal and hence partially match, or fail to mismatch, the lexical representation of *green*.

The underspecification of coronals is beneficial in dealing with variation, because coronals are most likely to undergo changes in connected speech. In the sentence "a quick rum picks you up", the word "rum" could mean *rum* or *run*, because the word-final coronal nasal in *run* can be pronounced as an [m], because it is followed by *picks* with an initial labial stop /p/ (Gaskell & Snoeren, 2008; Nolan, 1992). Having no specification for place of articulation of the coronal place of articulation is useful here, because the input "rum" with a labial nasal then does not mismatch the lexical representation of *run*. Even though the listener still has to decide—based on discourse information—whether the intended target was *run* or *rum*, phonological underspecification of coronals allows the recognition of the intended word. The value of underspecification becomes especially clear if we consider the cases in which the assimilation does not lead to confusion with another lexical item, as in the phrase "greem paper". Because "greem" matches (or fails to mismatch) the lexical representation of *green*, word recognition is straightforward in this case. Just as in the case of the memory benefit, not specifying the coronal place of articulation is again the most useful option, because intended labials are much less likely to undergo changes in connected speech.

The assumption of underspecification contrasts nevertheless with recent trends in (visual) perception research (Geisler & Kersten, 2002; Yuille & Kersten, 2006). This is not irrelevant as visual object recognition and color perception face similar problems as speech perception (Mitterer & de Ruiter, 2008). In all these domains, different distal stimuli can give rise to very similar proximal stimuli, and the same distal stimulus can—in different contexts—give rise to rather different proximal stimuli.

In the domain of visual perception, Bayesian statistical models have been proposed that assume that the perceiver performs optimally in combining prior knowledge and the perceptual input to calculate the posterior probability of the presence of an object in the input. In speech perception, this would mean to calculate the likelihood of a word being uttered given a certain input (i.e., p(Word/Evidence)). Recently, a model of spoken-word recognition has indeed been proposed based on these principles (Norris & McQueen, 2008).[2]

The differences between an optimal-perception account and the FUL model become clear if one considers how the models react to input phrases such as "gun production", "gum production", and "gum sales". The formula described for activation in the FUL model (Lahiri & Reetz, 2002), predicts the same level of activation for the word *'gun*' in all three phrases. However, an optimal-perception account would predict the following inequality relation: p(*'gun*'/"gun production") > p(*'gun*'/"gum production") > p(*'gun*'/"gum sales"). Why is this the case? First, even though *'gun*' can be pronounced as "gum", it is sometimes produced as "gun". And because "gun" is unlikely to be uttered when another word is intended, it follows that the likelihood of the intended word '*gun*' is higher if the input is "gun" than if the input is "gum". Secondly, '*gun*' is produced as "gum" only if the next word starts with a labial segment, hence the likelihood of the intended word '*gun*' is higher in the context "gum production" than in the context "gum sales".

While a few studies failed to find such a context effect (Wheeldon & Waksler, 2004), and it has been argued that this is evidence for an underspecification account (cf. Friedrich, Lahiri, & Eulitz, 2008), the majority of evidence seems to point towards the reality of such a context effect. At least five independent labs have now reported context effects, with a large variety of methods ranging from word monitoring to EEG and MEG measurements (Coenen, Zwitserlood, & Bölte, 2001; Darcy & Kügler, 2007; Darcy, Peperkamp, & Dupoux, 2007; Gaskell & Marslen-Wilson, 1996, 1998; Gaskell & Snoeren, 2008; Gow, 2002, 2003; Mitterer & Blomert, 2003; Mitterer, Csépe, & Blomert, 2006; Mitterer, Csépe, Honbolygo, & Blomert, 2006; Tavabi, Elling, Dobel, Pantev, & Zwitserlood, 2009).

The main focus of the current paper, however, is the assumption of mismatch tolerance if the input carries a deviant feature. The empirical tests here usually compared an assumed "no-mismatch" relation ('*lean*'/"leam") with an assumed mismatch relation ('*flame*'/"flane"). Note that the underspecification theory assumes that all features are extracted from the input. That is [Coronal] is extracted from the input and matched against lexical representations, even though [Coronal] is not specified in the lexicon. This leads to a "no-mismatch" relation between input and lexical

representation if an extracted non-coronal feature is mapped onto an unspecified place feature, but to a mismatch relation of a coronal feature in the input is mapped onto a lexically specified non-coronal place. There are several empirical findings that support this asymmetric matching assumption of underspecification. Lahiri and Reetz (2002) report a cross-modal priming study in which the auditory primes slightly mismatched the visual targets. In this study, auditory primes that mismatched their target on an unspecified feature ("greem" → *green*) primed lexical decisions more strongly than auditory primes which mismatched their target in a specified feature ("flane" → *flame*).

Another way to test the predictions of the FUL model about different types of mismatches is the electrophysiological component called Mismatch Negativity (MMN). This component arises if participants listen (passively or actively) to a train of stimuli, in which one type of sound is common and a small minority of the sounds is different from the rest (Näätänen, 1992, 1995; Näätänen, Pakarinena, Rinnea, & Takegata, 2004; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001; Schröger, Bendixen, Trujillo-Barreto, & Roeber, 2007). The infrequent type of stimulus elicits, in comparison to the standard stimulus, a negativity that is maximal around frontal midline electrodes, which is called the MMN. Eulitz and Lahiri (2004) used the vowels /o/ and /ø/ in such a paradigm, using both as standards and deviants. Importantly, /o/ is assumed to be specified as [Dorsal], while the coronal /ø/ is unspecified for place of articulation. Assuming that the standard stimulus is represented in a similar fashion to the lexical entry and that all features—even lexically unspecified ones—are extracted from the input, the FUL model predicts an asymmetry in the MMN depending on which stimulus is the standard. If the /o/—specified as [Dorsal]—is the standard, the deviant /ø/ provides the input feature [Coronal], which mismatches the representation of the standard. If, however, /ø/ is the standard, its coronal feature is not specified. Accordingly, the deviant /o/ (with the input feature [Dorsal]) leads to a "no-mismatch" relation between input and internal representation. In line with these predictions, the MMN was smaller and had a later peak latency when /o/ was the standard and /ø/ the deviant than when these roles were reversed.

However, not all empirical tests revealed such positive results. Gow (2001) used exactly the same design for a priming study as reported in Lahiri and Reetz (2002), but found perfectly symmetric amounts of priming. Marslen-Wilson, Nix, and Gaskell (1995) also found no evidence that the amount of cross-modal priming with slightly mismatching primes depends on whether the changed features are supposed to be specified or not. Tavabi, Elling, Dobel, Pantev, and Zwitserlood (2009) used the same MMN design as Eulitz and Lahiri (2004), and found no asymmetry depending on the specification/underspecification of the standard versus the deviant. They used nasals, thereby replicating an earlier negative finding (Mitterer, 2003). With regard to the MMN, an interesting alternative interpretation is suggested by a finding reported in Bonte and colleagues (Bonte, Mitterer, Zellagui, Poelmans, & Blomert, 2005). In one experiment, they used the Dutch pseudowords [nɔtfɛl] and [nɔtsɛl] both as standards and deviants in an MMN design. Note that these stimuli differ only in the fourth phoneme, and that the /s/ in [nɔtsɛl] is assumed to be unspecified for place of articulation,

while the /f/ in [nɔtfɛl] is specified as [Labial]. Each stimulus was used once as the standard with the other as deviant. Bonte et al. found a larger MMN with [nɔtsɛl] as the deviant than with [nɔtsɛl] as deviant. This would be in line with the prediction of an underspecification theory. The authors, however, had picked these stimuli because [nɔtfɛl] is more word-like (i.e., has a higher phonotactic probability in Dutch) than [nɔtfɛl]. This means that the asymmetry of the MMN could be attributed to the missing specification for place of articulation of the /s/ or to the higher phonotactic probability of the n-grams in [nɔtsɛl] as compared to [nɔtfɛl]. To disambiguate what caused the difference in MMN amplitude, Bonte et al. performed another experiment with the stimuli [so] and [fo]. The theory of phonological underspecification still predicts a larger MMN with the [so]-deviant, because its respective standard [fo] is specified with a labial place of articulation. These simple CV stimuli had a similar phonotactic probability, so that a frequency account predicts no difference in the MMNs. The results revealed equivalent MMNs independent of whether [so] or [fo] was the deviant. Therefore, the results of Bonte et al. show that the size of the MMN reflects the phonotactic probability of that sequence, or stated otherwise, the familiarity of the participants with a phoneme sequence (for the role of familiarity outside the language domain, see Jacobsen, Schröger, Winkler, & Horvath, 2005).

If prelexical frequency biases, such as phonotactic probability, can influence the MMN, it is not unlikely that the asymmetric MMN found by Eulitz and Lahiri (2004) reflects the higher frequency of coronal vs. non-coronal segments, rather than the underspecification of coronal segments.

More recently, Friedrich and colleagues (Friedrich, Lahiri, & Eulitz, 2006; Friedrich, et al., 2008) used later event-related potentials to test the prediction of asymmetric matching due to underspecification. They found that the pseudoword N400, an increased negativity for pseudowords in comparison to existing words, is smaller if the pseudowords mismatched real words only in an unspecified feature. That is, the N400 is smaller for pseudowords such as "greem", which does not mismatch the lexical representation of *green*, than to pseudwords such as "flane", which mismatches the lexical representation of *flame*. Secondly, they used fragment priming with mismatching primes, again comparing mismatches in specified and unspecified features (specified target words: "paprika" with "ta" as prime, underspecified target words: "tomato" with "po" as prime), and found more priming for underspecified "no-mismatch" targets in a component they called the "P350".

While it seems that these results may settle the issue, there remain two caveats, one statistical and one theoretical. The statistical caveat is that the EEG studies by Friedrich et al. (2006, 2008) involve a comparison between different words that are nested under the factor "specification". That is, a word such as *tomato*, which starts with a segment that is not specified for place of articulation, can obviously not be used in the "specified" condition because words in this condition have to start with a labial or a velar consonant. Especially in these circumstances, it is necessary to show that the results generalize over items (Clark, 1973), but the authors—following the tradition in psychophysiology—provide only a statistical test that shows the same result would be replicated with a new set of

participants, but no by-item analysis that would show that their result would be replicated with a different set of items.

An advantage of such EEG measures is nevertheless that they provide a window on earlier stages of processing than behavioral measures can. A disadvantage arises, however, with the interpretation of components. In the case of the studies of Friedrich et al. (2006, in press), the interpretation hinges on the assumption that the electrophysiological components reflect the ease of lexical access. For most electrophysiological components with a latency above 300 ms, the interpretation is less than straightforward and remains a matter of debate (see, e.g., Deacon, Hewitt, Yang, & Masanouri, 2000; Van Herten, Kolk, & Chwilla, 2005, for the N400 and the P600 components, respectively). It is hence also quite possible that the P350 and the N400 effects in the data of Friedrich et al. reflect pre-lexical processes, which react differently to high- and low-frequency phonemes or phoneme sequences, or that they reflect post-lexical processes related to the meta-linguistic lexical-decision task that participants had to perform in these experiments.

In the light of these considerations, a more stringent test would be possible with a measure that also provides a window on the early stages of lexical access, but is more strongly related to activation of lexical items. Over the last decade, eye-tracking in a visual-world paradigm proved to be a useful technique for such purposes. Allopenna, Magnuson, and Tanenhaus (1998) presented participants with pictures of four objects on a screen and a spoken instruction to move one of the objects (e.g., "pick up the beaker"). The four pictures not only contained the target object (a picture of a beaker) but also two pictures of objects with similar names to the target. One object had a name with an onset overlap with the target (a beetle) and one a rhyme overlap (a speaker). Lexical activation of such competitors is evaluated by comparing looks to these objects with looks to an unrelated distractor object (a carriage). Results showed that participants looked to the target and the onset-overlap competitor more than to the unrelated distractor, with looks to both starting 200 ms after word onset. There initial looks showed no difference between target and onset-overlap competitor. While this may be surprising, it is important to consider here that planning and executing an eye-movement takes about 150-200 ms (Hallett, 1986). Therefore, fixations 200 ms after word onset are most likely to be based on the initial segment of a word alone. This explains why such early fixations revealed no preference for the actual target over the onset-overlap competitor. They are based on information (e.g. the initial syllable "bi") which does not distinguish between target and competitor (e.g. "beetle" and " beaker"). These fast early signal-related looks demonstrate that eye-tracking provides a window on the earliest stages of lexical processing.

This method matches the advantage of ERPs of delivering an on-line, time-locked measure of cognitive processes. However, the chain of inference is considerable smaller. If a picture attracts more looks than an unrelated distractor, it provides clear evidence for lexical activation of the word related to the picture. However, a remaining disadvantage of the design as used by Allopenna et al. is that one has to find phonologically related word pair that are picturable. To overcome this problem, McQueen

and Viebahn (2007) used printed words instead of pictures and found essentially the same results as Allopenna et al. (1998). The visual-word paradigm with printed words was therefore used here to compare the predictions about lexical matching made by the FUL model with those of an optimal-perception account.

Experiment 1

In this experiment, Dutch minimal pairs were used, in which one member started with a coronal segment and the other with a labial segment (e.g., *mier-nier*, Engl. 'ant' and 'kidney'). We used only onset minimal pairs because phonological underspecification is position independent and effects in eye-tracking are particularly strong for onset overlaps and mismatches.

Participants heard an instruction to click on one word ("Klik op het woordje mier", Engl. 'click on the word *ant*') and the visual display contained the written target word (*mier*) and its competitor (*nier*) and two unrelated distractors. One potential problem with such a design is that it may force participants to listen quite carefully. Only a short lapse in attention is sufficient to make the distinction between *mier* and *nier* difficult. To make the task less taxing, we made use of a recent innovation of the visual world paradigm. Mitterer and McQueen (2009) presented printed words with accompanying objects, which makes the instructions easier to follow. Figure 1 shows an example of such an experimental display. If a participant hears the instruction "Klik op het woordje mier above the star" ('Click on the word *ant* above the star'), he or she does not need to listen carefully to hear the difference between [mir] and [nir], because the phonologically distinct object name will (redundantly) specify the printed word *mier* as the target word, because it is the one above a star, while *nier* is above a triangle.

In such a paradigm, the FUL model predicts that the spoken target [mir] (for the printed word *mier*) should lead to many looks to the phonologically similar printed word *nier*, because the input [mir] matches the underspecified lexical representation of *nier*, which is /[nasal]ir/. The spoken target [nir] (for the printed word *nier*), however, should lead to a much smaller amount of looks to the phonologically similar, printed *mier*, because the input [nir] mismatches the specified lexical representation of *mier*, which is /mir/.

The optimal-perception account, in contrast, predicts a symmetric pattern of competition. The looks to the printed word *nier*, given the auditory input [mir], should occur just as often as looks to the printed word *mier*, given the auditory input [nir]. A reviewer suggested that the optimal-perception account should predict asymmetric competition as well, because /n/ can sometimes be pronounced as an [m] (in case of place assimilation). However, place assimilations only occur in word-final position. In the current experiment, the nasals occur in word initial position, and are preceded by a schwa. The critical a-priori probabilities p(*nier*/[mir]) and p(*mier*/[nier]) provided by the production patterns of Dutch are therefore so small that they can hardly influence the eye movements.

Following the procedure of Mitterer and McQueen (2009), there were also "semantic" filler trials in which the same printed word appeared twice on the screen. In this case, there are no phonological cues to what the target is, and the target is identified by adding the second part of the instruction ("...above the star"). Disambiguation of the target was achieved either by the preposition (because one printed word was above and one next to a star) or by the object name (because one printed word was above a triangle and one above a star). These trials were included because Mitterer and McQueen found that phonological effects were more robust when such semantic trials were included in the experiment.

Method

Participants

Twenty-four native speakers of Dutch from the Max-Planck Institute's participant pool participated in the experiment for pay. There were no problems with the calibration of the eye-tracking, so that data from all participants could be used.

Stimuli

Sixty Dutch minimal pairs were selected, with either coronal or labial initial consonants. There were 20 pairs each with nasals (e.g., *mier-nier*, Engl. 'ant' and 'kidney'), fricatives (*fabel-sabel*, Engl. 'fairy tale' and 'saber'), and voiced stops (*bom-dom*, Engl. 'bomb' and 'studip'). These were used to test the differential predictions of the FUL model and an optimal-perception account. The Appendix lists these items. Additionally, we selected 20 minimal pairs in which one member had an initial /t/ and the other either a /p/ or a /k/ (e.g., *tonijn* - *konijn*, Engl. 'tuna'-'rabbit'). These were used in the semantic trials, in which either the object name or the preposition (i.e., a word is either *above* or *next to* an object) disambiguated what the target object was.

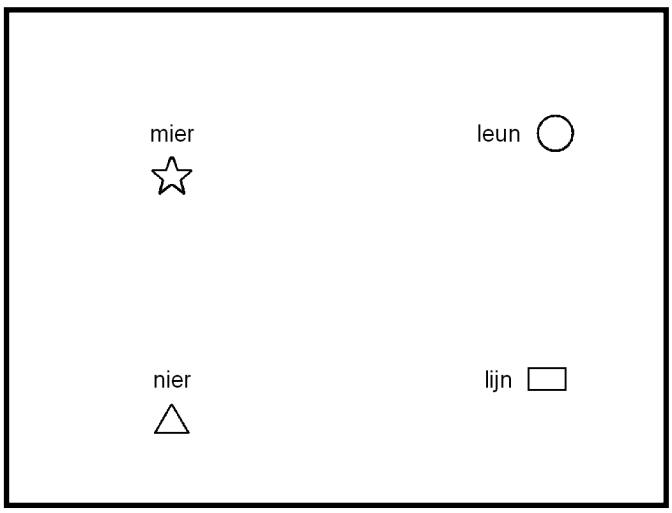

Figure 1. An experimental display as used in Experiment 1 with the instruction "Klik op het woord mier above the star" ('Click on the word ant above the star').

The items with labial and coronal onsets had similar lexical frequencies (log, per million; labial: 1.02, coronal: 0.82, $t(118) = 1.3$, $p > 0.1$). All but six words (three labial and three coronal) were not unique until their offset. Given the importance of the onset information, we also checked how many words shared the same onset and vowel. Again, this was roughly comparable for the (log) number of words (labial: 5.01, coronal: 5.2, $t(118) = 0.9$) and their frequency (labial: 9.7, coronal: 10.3, $t(118) = -1.7$, $p = 0.1$). Because these pairs are not perfectly matched, we will use these lexical measures as co-variates in the data analysis. This procedure renders the design more powerful than picking only a subset of the items that would be slightly better matched (Baayen, 2008).

The minimal pairs were spoken in a sentence context "klik op het woordje ... [boven/naast] de [rechthoek/driehoek/cirkel/ster]" by a female native speaker of Dutch with no knowledge of the objective of the experiment. Each word was read twice, once with the preposition *boven* (above) and once with the preposition *naast* (next to) and a random object name. The speaker was instructed to read the sentences with a small gap between the target word and the following preposition. The experimental stimuli were spliced together from two parts. Splicing was necessary because recording every word in every possible combination of preposition and target object would have made the reading list prohibitively long. The first part contained the pre-target context and the target itself ("klik op het woordje *target*"), and the second part the preposition and the object name. To select fragments, the mean pitch was calculated for each sentence and the grand mean was calculated from that. Then one sentence of each combination of the two prepositions (boven & naast) and four objects (rechthoek, driehoek, cirkel, ster) with a mean pitch close to the grand mean pitch was extracted to be used as post-target context. For the initial part of the instruction used during the experiment, the first part of the experimental sentence was then extracted from the one sentence with the mean pitch closer to the grand mean (except for the eight sentences used for the second-part fragments, so that all materials were cross-spliced). In this way, we created the 1280 sound stimuli used in this experiment (160 words * 2 prepositions * 4 objects).

The visual stimuli with the printed words were generated as bitmaps in the font Arial with (pointsize: 28), which always had a height of 42 pixels and varied in width from 43 pixels to 167 pixels, depending on word length and the width of the letters ("m" taking up more space than "l"). The bitmaps were positioned with their center point on the centers of the four quadrants of the screen (see Figure 1). The bitmaps with the objects had a size of 70 by 70 pixels and were positioned directly under or directly right of the bitmaps with the printed words, so the distance between the word and the object was independent of word length. For the semantic trials, the screen was filled with two times the target word and two times its phonological competitor. On ten trials, the word to be clicked on was disambiguated by the position; on the other ten trials it was disambiguated by the object. That is, with the instruction "klik op het woordje team boven de ster" ('Click on the word *team* above the star'), the display contained the printed word *team* twice. On same trial, the two identical printed words appeared

above and next to a star, so that the preposition "boven" ('above') disambiguated the instruction. In the other case the printed word team appeared both above a star and above a circle, so that only the object name disambiguated the instruction.

Apparatus and Procedure

Participants were seated in a chair in front of a computer screen and an EyeLink 1000 eye tracker from SR Research. This eye-tracker samples the position of one eye with a frequency of 1kHz. The right eye was selected to be tracked in all current experiments. The stimulus presentation was controlled by a different computer using the ExperimentBuilder software from SR Research. After calibration, written instructions (presented on the screen) told participants to follow the spoken instructions, which they would hear over headphones, to click on certain words. The pre-experimental, written instructions also mentioned that sometimes a word would appear twice on the computer screen, and the one to click on would be identified by a following phrase such as "above the star". Each experiment started with two of these semantic filler trials.

Each participant completed 80 trials; 60 trials in which the target word was accompanied by a competitor that differed only in the place of articulation of the initial consonant, and 10 filler trials each in which the same word appeared twice on the screen and the target was identified by either the position or the identity of the accompanying object. Hence, each participant heard only one member of each of the 60 minimal pairs and completed 30 trials each in which the target started with a specified labial or a supposedly underspecified coronal consonant. The allocation of target and competitors were counterbalanced across participants.

A different random order was generated for each participant. Because participants have a tendency to first look at the top left item, target and competitor positions were counterbalanced for each participant, so that the target and its competitor were presented 20 times each at each of the four screen positions: 15 times in the 60 phonological trials and 5 times in the 20 semantic filler trials. In addition, the target objects were counterbalanced so that the word to be clicked on was accompanied 15 times each in the phonological trials and 5 times each in the semantic trials by the star, the circle, the triangle, and the square.

Design and Analysis

The data analysis was based on the event-file of the eye-tracker, which categorizes the eye-position samples into fixation, saccades and blinks. For the data analysis, distance between the eye position and the four target objects on the screen was sampled at—on average—10ms intervals. No fixed sampling frequency was used to account for the different durations of the different sound files. The target words, for instance, were on average 600 ms long; hence the eye-position was sampled at 60 equal time steps based on the duration of the target word (e.g., steps of 6 ms if the target word on a given trial had a duration of 360 ms but steps of 15 ms if the target word had a duration of 900ms).

The dependent variable was—following Mitterer and McQueen (2009)—the Euclidean distance (in pixels) between the objects and the fixation position. This makes it superfluous to define an area of interest for which fixation count as being "on" a given object. The dependent variable was the distance to the competitor, that is, the word phonologically similar to the target, in a time window from 200 ms after target onset until the asymptote of the target distance function at 2000 ms after sentence onset (see Figure 2). In addition to the subject and item random effects and the fixed factor Underspecification (specified vs. unspecified place of articulation of the onset of the target word), the manner of articulation of the first phoneme and the lexical statistics were added as fixed factors. Lexical statistics were removed in a step-wise backward elimination if not significant.

Results

Performance was near ceiling with 100% correct responses for specified targets and 99.7% for underspecified targets (i.e., 2 errors on 750 trials). Figure 2 shows the distance between the different objects and the fixation positions on correct trials. There seems to be a small preference for underspecified competitors. The competitor with the underspecified onset (from the "specified target condition") received more looks than the competitor with the specified onset (from the "underspecified target condition) but this difference was not significant (B = 8.5, SE(B) = 5.9, t = 1.34, $p_{MCMC} > 0.1$, d = 0.10). When lexical statistics were included in the model, the only significant effect was the Cohort Frequency of the competitor ($p_{MCMC} < 0.05$, $R^2 = 0.004$), and the effect of underspecification remained insignificant (B = 10.8, SE(B) = 6.3, t = 1.70, $p_{MCMC} = 0.09$, d = 0.11) [3].

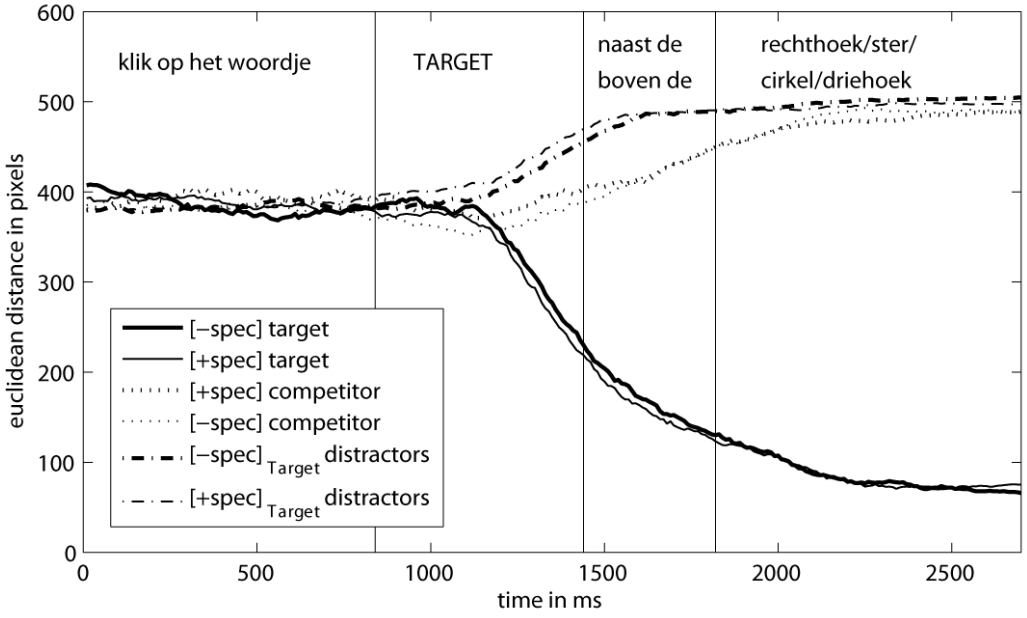

Figure 2. Eye-tracking results from Experiment 1. The vertical lines indicate the onset of the different parts of the instructions, which is given in an English translation. The condition labels (± spec) indicate target and competitor specification. Note that the lines for the specified target and the underspecified competitor are taken from the same trials, as specified targets necessarily go together with underspecified competitors (and vice versa).

Discussion

The results are in line with an optimal-perception account and contradict the predictions of an Underspecification account. Participants did not look more at words with onsets that were unspecified for place of articulation when hearing (a non-mismatching) specified onset than they looked at words with a specified onset when hearing (a mismatching) coronal onset. Nevertheless, we did observe a trend in the direction predicted by the Underspecification account. Might the lack of a significant effect simply be due to a lack of power? The statistical model reveals an unstandardized coefficient of 10.8 for the Underspecification factor, meaning that the model estimates that the mean distance between eye position and competitor is 10.8 pixels smaller if the competitor starts with a segment that is not specified for place of articulation. The standard error of this coefficient is 6.3 pixels, leading to a non-significant difference. To examine whether this standard error is excessively large and prevents us from finding an effect, it is useful to consider the effect sizes usually found in visual-world experiments testing phonological effects. Two papers examined context effects in the recognition of phonological variants with a similar eye-tracking paradigm (Gow & McMurray, 2007; Mitterer & McQueen, 2009). As those papers investigated rival theories to the FUL model, it is reasonable to assume that effects of phonological underspecification should be of a similar size as the effects reported in these experiments. Mitterer and McQueen found effect sizes between 17 and 33 pixels. Gow and McMurray found a difference in fixation proportion of 7%. Assuming that fixations have an average distance to the target centre of about 100 pixels and non-target fixation are evenly distributed to the other three objects on the screen, this translates into a difference in distance of 27 pixels. This seems to indicate that phonological effects in eye-tracking paradigms can be expected to be around 20 pixels. Such an effect would be significant in the current experiment given the current standard error of the regression coefficient (6.3).

Even though the lack of an effect is hence unlikely due to a lack of power, the lack of an effect might, however, be due to an experiment specific strategy induced by the minimal pairs. Friedrich et al. (2008) added the idea of "phonological parsing mechanism" to the model of phonological underspecification, which helps the model in taxing situations. Friedrich et al. used this mechanism to explain why participants overwhelmingly respond "no" when hearing "leam" in a lexical decision task. According to the FUL model, the input "leam" and "lean" match the lexical representation equally well, so that additional mechanisms are needed to explain why "lean" leads to "yes" responses and "leam" to "no" responses in a lexical decision task. Friedrich et al. argue that phonological parsing mechanisms are invoked in taxing situations. These mechanisms create the output form of words, adding the coronal place of articulation by a post-lexical rule. As this output form deviates from "leam", this phonological parsing mechanism allows participant to respond with "no" in a lexical decision task. Using minimal pairs may also invoke such special phonological processes, which mask

the effects of underspecification. Therefore, a second experiment was devised in which participants were not exposed to minimal pairs.

Experiment 2

In this experiment, we used words as target-competitor pairs which had an overlap in their initial phonemes, but were clearly distinguished by their offset, such as the Dutch word pair *kogel-toga* /koxəl/-/toxa/ (Engl. 'bullet' and 'academic gown'). If we take into account that, under the underspecification assumption, the lexical representation of /toxa/ does not specify a place of articulation for the first segment, the input [koxəl] matches up to the third phoneme. Because the lexical representation of the word /koxəl/ specifies that the first segment is velar, the input [toxa] produces a mismatch in the onset position, resulting in minimal lexical activation. Phonological underspecification then predicts that, in a visual word paradigm, the printed word *toga* given the auditory input [koxəl] should attract more looks than the printed word *kogel* given the auditory input [toxa]. In contrast, the optimal-perception account predicts that competition should be symmetric.

Method

Participants

The same 24 participants who took part in Experiment 1 also participated in this experiment. Both experiments were run in one session. To avoid an influence of seeing minimal pairs during this experiment, this experiment was in fact run as the first in the session. As in the first experiment, calibration was successful for all participants.

Stimuli

48 word pairs with some phonological overlap after the first segment (such as *toga-kogel*, in phonological transcription /toxa/-/koxəl/) were used (see the Appendix). The initial segment was a nasal in six pairs, a voiceless stop in seven pairs, a voiced stop in 18 pairs, and a fricative in 17 pairs. Disregarding the onset mismatch, the mean overlap was 3.3 segments between these pairs (see the Appendix). For each pair, two additional words were used as distractors that had no clear phonological overlap with the word pair. Eight additional quadruplets of phonologically unrelated words were selected to serve as items on eight practice trials.

The two groups of items—with specified and unspecified place of articulation in the onset position—were reasonably matched in their lexical statistics (log frequency: 0.4 vs. 0.1, log cohort size: 2.5 vs. 2.4, log cohort frequency: 2.9 vs. 3.1, uniqueness point: 5.9 vs. 5.4, $t_{max}(94) = 1.5$, $p_{min} > 0.1$). To deal with the remaining differences, these measures were taken into account in the data analysis.

All target words were produced in a sentence frame "klik op het woord ..." ('click on the word ...') by a female speaker of Dutch. The uncut recordings were used as experimental stimuli. As in Experiment 1, the visual stimuli were generated in the Arial font (pointsize: 28 size). The resulting

bitmaps always had a height of 42 pixels and varied in width from 157 pixels for the word *krijt* (Engl. 'chalk') to 244 pixels for the word *meegaand* (Engl. 'compliant'). The bitmaps were positioned with their center point on the centers of the four quadrants of the screen.

Apparatus and Procedure

The Apparatus was the same as in Experiment 1. Each participant completed 56 trials starting with 8 practice trials that were identical for all participants. Each participant then heard a random order of 48 trials. Of these, 24 trials had a target word that started with a coronal consonant and 24 trials had a target word that started with a labial or velar consonant. A given participant heard only one member of a phonologically related pair as target, and the numbers of presentations of the coronal and non-coronal members of a pair were counterbalanced over participants. A different random order was generated for each participant, with the constraint that targets and competitors appeared equally often on the four screen positions.

Design and Analysis

The Design and Analysis was similar to Experiment 1, including the data reduction. The critical independent variable was the assumed phonological specification of the place of articulation in the onset position of the target word. As the dependent variable, the mean distance to the competitor in the window 200 ms to 600 ms after target onset was used. For the statistical analysis, a mixed-effect model was used. Subject and Item were entered as random factors, lexical statistics, manner of onset phoneme, and underspecification were entered as fixed factors. Lexical statistics were removed in a step-wise backward elimination if not significant.

Results

The click responses were highly accurate with only 3 errors over 1200 experimental trials. Figure 3 shows the eye-tracking data for the correct trials, revealing no clear difference in competitor activation. Note that the assumption of phonological underspecification predicted that competitors with an onset that is unspecified for place of articulation should be activated more strongly than competitors with an onset that is specified for place. However, there is no such apparent difference in Figure 3. In line with what Figure 3 suggests, the statistical analysis revealed no difference between the two competitors in both conditions (B = 4.3, SE(B) = 9.7, t = 0.45, $p_{MCMC} > 0.5$). Adding item properties revealed an additional effect of manner of articulation of the first segment. Participants looked more at the competitor on trials in which target and competitor were nasals ($B_{manner=nasal}$ = -45 pixels, t = -2.83, $p_{MCMC} < 0.01$, d = 0.27).

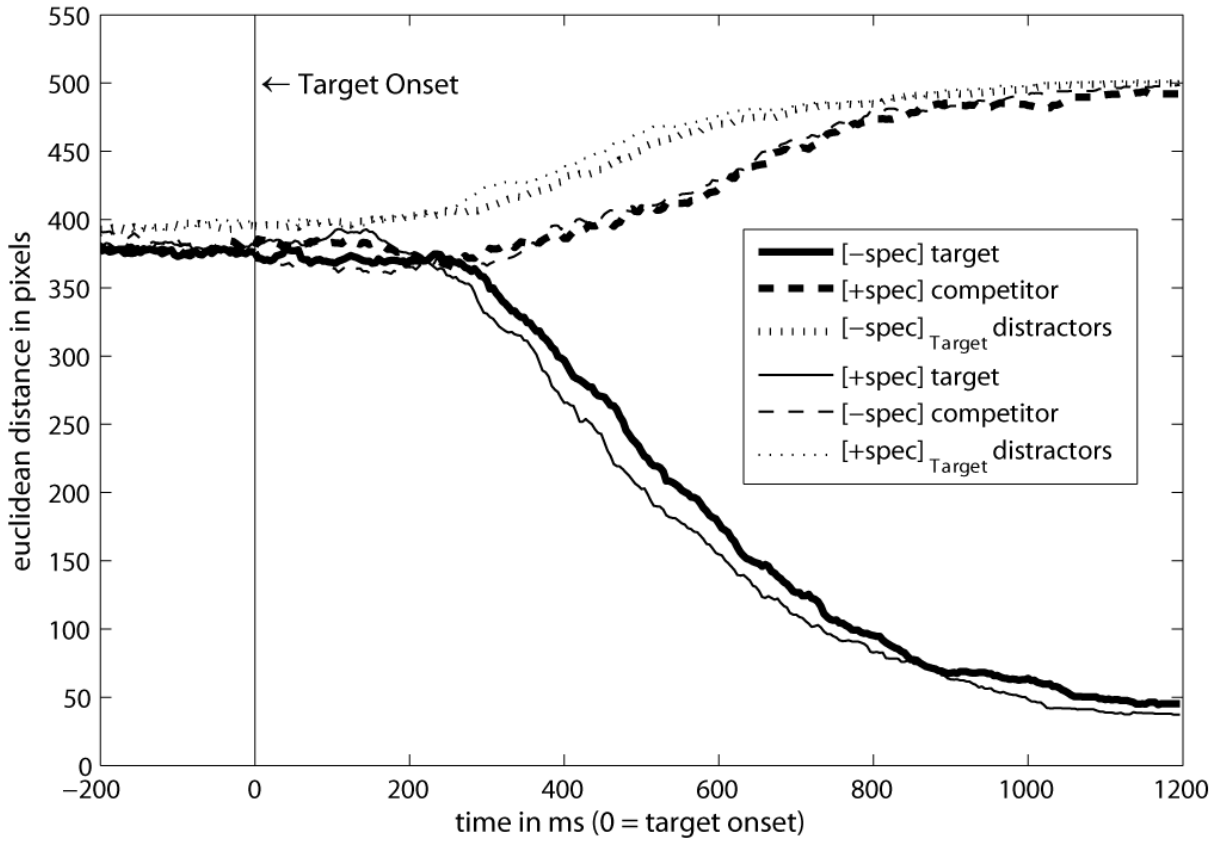

Figure 3. Eye-tracking results from Experiment 2. The condition labels (± spec) refer to the specification of target and competitor (see also Figure 2).

Discussion

The results are in line with those of Experiment 1. The lexical activation of words not matching the input is independent of the place of articulation of the mismatch. The analysis revealed an effect of manner of articulation, with most looks to competitors with a nasal onset. This is not surprising, given that nasals are difficult to distinguish in terms of place. This is especially true for the first part of the signal, which only contains the nasal murmur, which, especially in onset position, contributes only very little to place perception (Repp & Svastikula, 1988).

The results are in line with the assumption of an optimal-perception account. Just as in visual perception, spoken-word recognition seems to use the incoming information in an optimal way. The results therefore contrast with the view that the lexical representations are underspecified, which would mean that some information would not be used. Instead, we find that a mismatch of labial or velar input with a coronal onset leads to a deactivation of the target just as strongly as the mismatch of coronal input with a labial or velar onset.

Nevertheless, there is a possibility to explain the current results without questioning the assumption of underspecification. The input [kox] matches all specified features of the words *kogel* and *toga*. Nevertheless, *kogel* will be activated more strongly, because lexical activations in the FUL model also take into account the number of matching features (see Formula [3] in Lahiri & Reetz,

2002), which is larger because *kogel* matches the initial place of the input. In simpler terms, the FUL model takes into account that a "match" relation is better than a "no mismatch" relation. If the eye-movements in a visual-world experiment are further influenced by a "winner-takes-all" mechanism, looks to *toga* will be inhibited by the presence of the better-matching *kogel* in the display. In other words, the presence of a matching target on the input screen inhibits looks to the underspecified competitor. This alternative account would, however, predict that the effect of asymmetric matching between input and mental lexicon should arise if the matching target is not on the screen. This was tested in Experiment 3.

Experiment 3

In this experiment, we further explored the contrasting predictions of an optimal-perception account and the assumption of phonological underspecification. We used the version of the visual-world paradigm in which the auditory input corresponded to none of the printed words on the screen. Huettig and Altmann (2005) showed that the absence of a perfectly matching target on the screen increases the amount of looks to related objects. A design without targets might hence be more sensitive to reveal an asymmetric effect of lexical matching.

The items on the critical experimental trials were the same as in the previous experiment. What changed with respect to the previous experiment is that the word that appeared in the instruction sentence was replaced on the screen by a phonologically unrelated distractor. Participant heard an instruction such as "klik op het woordje kogel" (Engl. 'click on the word bullet'), and saw on the screen the printed words *toga*, *zucht*, *draak*, and *wreed* (Engl., 'academic gown', 'sigh', 'dragon', and 'cruel'). The word *toga* is the only word with a phonological relation to the target, having the same vowel in the first syllable and same onset in the second syllable. In addition, to reiterate the assumption of phonological underspecification, the initial /t/ is supposed to be unspecified for place of articulation. Hence it does not mismatch the initial [k] of the auditory input provided by *kogel*, and therefore provides a *perfect* match for the spoken input: As a consequence, the input [tox] should not activate *toga* more strongly than the input [kox].

Method

Participants

24 native speakers of Dutch from the Max Planck subject pool took part in this experiment. None of them had participated in Experiments 1 or 2. There were no problems with the calibration of the eye-tracking, so that data from all participants could be used.

Stimuli, Apparatus, and Procedure

The auditory stimuli for the critical experimental trials were the same as in Experiment 2. The visual displays were changed so that the printed word corresponding to the spoken word was replaced by a phonologically unrelated word. Thus, participants heard a sentence such as "klik op het woord

toga" and saw on the screen the four printed words *kogel*, *zucht*, *draak*, and *wreed*. Because participants have nothing to click on, they were instructed to click on the middle of the screen if they found no matching target.

To encourage participants to look for a matching target, we added 96 filler trials, in which there was a matching target on the screen, so that a word had to be clicked in 2/3 of the trials. The experimental displays for these filler trials contained, like the experimental trials, four phonologically unrelated printed words, one of which corresponded to the spoken target.

The apparatus was the same as in the previous experiments. As in Experiment 2, a different random order was generated for each participant, counterbalancing the positions of targets and pseudotargets for each participant. These orders contained at least five filler trials at the start.

Design and Analysis

Design and Analysis were the same as in Experiment 2.

Results

Participants again performed the task very accurately. In the data set of 1200 experimental trials, there was only one error in which a participant clicked on a competitor, while there were only 51 filler trials (of 2400, hence 2.1%) in which the correct target was not clicked on.

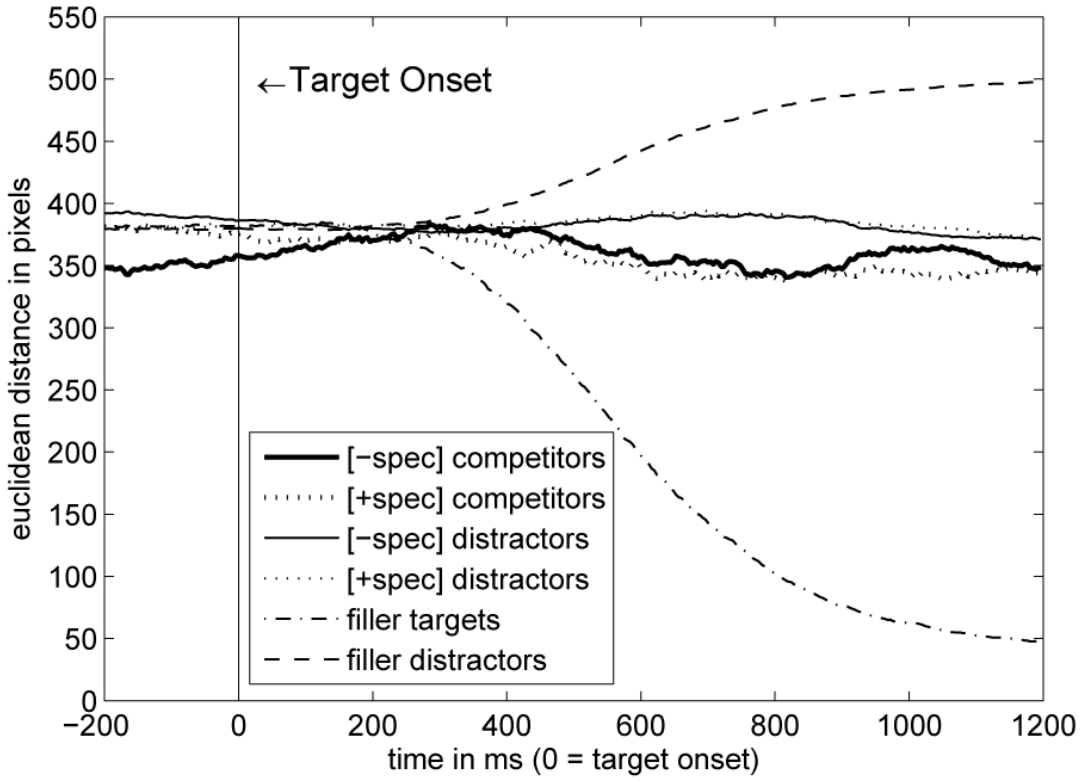

Figure 4. Eye-tracking results from Experiment 3. The critical comparison is between the specified and underspecified competitor, which do not differ.

Figure 4 shows the eye-tracking data for the experimental trials with correct responses, with the filler data as a reference frame. The competitors got less looks than the matching targets on the filler trials, but still more looks than distractors on the experimental trials. Comparing the two types of competitors, participants looked more at the competitors with an onset that is specified for place, but this difference is far from significant (B = -8.4, SE(B) = 9.6, t = -0.86 $p_{MCMC} > 0.2$). Note that the direction of this difference is in the opposite direction of what an underspecification account predicts. The regression analysis with the lexical statistics revealed no additional effects. Participant did, however, look more at competitors starting with stops than at competitors starting with a fricative (t = -2.3, $p_{MCMC} < 0.05$, d = 0.15), while the preference for nasals over fricatives was only a trend (t = -1.6, $p_{MCMC} = 0.10$, d = 0.16).

Discussion

This experiment replicated the main result of Experiment 2. Given a mismatch in place of articulation in the onset position, participants did not look more at words with a coronal onset—which is assumed to be unspecified in the FUL model—than at words that have a specified place in the onset position. Moreover, the time course data are rather telling. In all three experiments, the results show a rather sharp drop in distance to the target around 250 ms after target onset. According to the theory of phonological underspecification, the activation level of underspecified words is identical for coronal vs. labial or velar inputs. The input features labial or velar, as well as the "correct" coronal feature, constitute a "no mismatch" with the lexical representation of a word that has segments with no specified place of articulation. Differences might arise due to lexical inhibition from other words, but those again need time to build up (Pitt & Samuel, 2006). Therefore, the initial activation for underspecified words should not differ as a function of the exact input. In contrast with this prediction, we see a sharp decrease in the distances between eye positions and targets around 250 to 300 ms, but never for mismatching competitors, independent of their specification. While the absence of such a pattern may be explained by the presence of a matching target for the underspecified competitor in Experiment 2, this is not the case in the current experiment.

Experiments 1 to 3 thus failed to provide evidence for the assumption that the lexical representation of words with coronal consonants is not specified for place of articulation in these three experiments. While there was a trend in the predicted direction in Experiment 1, an opposite trend was observed in this experiment. Moreover, in all three experiments, the standard errors for the regression weights of the underspecification factor were small enough (≤ 10) to give rise to significant effects, taking typical effect sizes in similar experiments (± 20 pixels) as a reference point. Overall, the results indicate that the matching between input and lexicon treats all mismatches alike. This pattern is in line with an optimal-perception account.

In all cases, the optimal-perception account predicted a null-effect and the Underspecification theory predicted the presence of an effect. Given that current statistical practices do not allow us to

accept the null-hypothesis, the presence of similar effects in other previous experiments, as cited in the introduction, may, however, be seen as more important than the current series of null results. Such an argument can be best tackled by designing an experiment in which the assumption of Underspecification predicts a null result, while the optimal-perception account predicts a difference between conditions. This was the purpose of the fourth and last experiment.

Experiment 4

While the empirical evidence and counterevidence has mostly focused on the underspecification of place of articulation, other features are also assumed to have specified and unspecified feature values. For Dutch, the assumption is that [+voice] is lexically specified and [-voice] is the default feature that is not lexically specified (van der Feest, 2007, see p. 34). Accordingly, the representation of /t/-initial words in Dutch (such as /tart/, 'cake') only specifies that the word starts with a stop. Accordingly, other inputs match independently of place of articulation (such as the word /part/, 'horse'), voicing (such as the word /dart/, 'dart'), and the combination thereof (such as the word /bart/, 'beard'). It follows that the Dutch word *taart* /tart/ should be equally strongly activated when hearing the word *paard* /part/, with only a mismatching place of articulation, as when hearing the word *baard* /bart/ with a mismatching place of articulation and mismatching consonant voicing.

This prediction follows directly from the formula (3) for lexical activation in the FUL model presented in Lahiri and Reetz (2002). According to this formula, the activation level of a word is determined by the (squared) number of matching features divided the product of the number of features extracted from the input and the number of features in the lexicon. What does this mean for the activation of the word /tart/ given the inputs [part] or [bart]? Clearly, the number of features in the lexicon is the same in both cases (all the specified features for the word /tart/). Moreover, the number of features specified in the input is identical as well because [+voice] is extracted for [bart] and [-voice] for [part], all else being equal. That is, whether a feature is lexically specified or not, it is always extracted from the input. Finally, the number of matching features is the same as well, because both the extracted [+voice] for [bart] and [-voice] for [part] produce a no-mismatch with the lexical representation of /tart/, which has no specification for voicing in the onset position. In short, there is no difference here because the number of "no mismatch" relations between input and lexicon is not used in the scoring formula presented by Lahiri and Reetz (2001, p. 641).

However, if one assumes that the mental lexicon makes use of all cues—that is, it behaves optimally—, one has to predict that the one-feature mismatch (*paard*, /part/) should lead to a stronger activation of *taart* than the two-feature mismatch (*baard*, /bart/). Note that the roles of predicting an effect vs. predicting a null-effect are now reversed. The assumption of phonological underspecification predicts a null-effect, and the optimal-perception account predicts a difference between conditions. Previous research indicates that it should be possible to find a difference between a one- and a two-feature mismatch. The activation of a word that does not match the input perfectly is influenced by the

number (Connine, Blasko, & Titone, 1993) and even the type of features (Ernestus & Mak, 2004) that mismatch.

This experiment again involves minimal pairs. As in Experiment 1, we therefore used carrier sentences with redundant information to identify the target object. This was again achieved with instructions such as "klik op het woordje paard boven de ster" ('click on the word horse above the star'). Because the display were arranged in such a way that the printed word *paard* was the only one above a star, participants were not forced to listen extremely attentively to determine that the target word was *paard* /part/ and not the confusable competitor *taart* /tart/ ('cake'). Just as in Experiment 1, we also included "semantic" trials in which the target to click on was disambiguated by either preposition or the object name.

Method

Participants

The same 24 participants that took part in Experiment 3 also took part in this experiment. It was run after Experiment 3 in one experimental session.

Stimuli

40 /t/-initial words were selected that had one phonological neighbor that differed only in place of articulation (24 starting with a /k/ and 16 starting with a /p/) and another phonological neighbor that differed in place of articulation and voicing (necessarily starting with /b/, because the voiced velar fricative /g/ only occurs in a few loan words in Dutch). These were spoken by a female speaker of Dutch in sentence frames such as "klik op het woordje ... boven de ster" ('click on the word ... above the star'). As in Experiment 1, the sentence frame varied, and could contain the preposition *naast* ('next to') instead of *boven* and instead of the object name *ster*, the object names *cirkel* ('circle'), *rechthoek* ('rectangle'), or *driehoek* ('triangle'). Each target word was recorded in two instruction sentences, one with the preposition *boven* and one with the preposition *naast* and a randomly determined object name following it. As in Experiment 1, the instructions used during the experiment were concatenated using a sound-file for the part up to and including the target words, and a second part containing the preposition and the object name. Selection was, as in Experiment 1, based on the mean pitch of the recorded utterances.

The target words for /t/-targets, unvoiced stop targets, and voiced stop targets were similar in their lexical statistics, such as frequency (/t/: 1.96, unvoiced stops: 2.39, voiced stops: 2.23, in log10 per million), the number of words starting with the same onset and vowel (/t/ 2.23, unvoiced stops: 2.35, voiced stops: 2.22, in log10), and the frequency of these onsets (/t/, 2.60, unvoiced stops: 2.54, voiced stops: 2.39, in log10 per million), and their uniqueness points (/t/ 4.75, unvoiced stops: 4.48, voiced stops: 4.28). These lexical statistics were again taken into account in the regression model.

The printed words were, as in the previous experiments, generated as bitmaps using the Arial font (28 pointsize). This results in bitmaps with a height of 42 pixels and a width varying from 43

pixels to 182 pixels. The printed words were placed on the quadrants of the center of the screen, and the objects (in bitmaps with 70 by 70 pixels) were placed directly under or directly to the right of the printed words.

Apparatus and Procedure

The apparatus was the same as the previous experiments. There were 120 trials for each participant. The most critical of those were forty trials on which a /t/-initial word (e.g. /tart/) was used as a competitor for a target that differed only in place of articulation (e.g., /part/) or in both place and voicing (e.g., /bart/). Given the forty minimal triplets, this means that a given participant saw the printed word *taart* only once, as a competitor for the word *paard* or *baard*. Over participants, each /t/-initial word appeared equally often as competitor for a word with a voiceless stop (*paard*) and as competitor for a word with a voiced stop (*baard*).

To prevent participants from ruling out /t/-initial words as targets, 40 trials were added in which a /t/-initial word was the target and the competitor differed in place only or in place and voice, on 20 trials each. This design also allowed another test of the basic assumption of asymmetric lexical matching. According to the FUL model, competitors with initial /t/ should receive more looks than competitors with initial /p/, /k/, or /b/.

A different random order was generated for each participant. As in the previous experiments, target and competitor locations were counterbalanced for each participant. As in Experiment 1, the lists also contained 40 filler trials in which participants saw two identical printed words on the screen and the target was determined by the preposition phrase (e.g. "above the star"). On twenty trials the preposition determined which printed word was the target (because one was above and one next to a star) and, on the other twenty, the object name determined which one the target was (because one was above a triangle and one above a star).

Design and Analysis

The design allows two tests that distinguish the FUL model from an optimal-perception account. First of all, the question is, as in the previous experiments, whether competitors with unspecified features in onset position (/t/ in this case) attract more looks than competitors with specified features in onset position. The second question is whether the competitors with unspecified values for place and voicing in the onset position (i.e., /t/) attract more looks if they deviate from the target in only the place feature than if they deviate in both place and voicing features. For both questions, the dependent variable was the mean distance to the competitor in a window 200-600 ms after target onset. Independent variables were the underspecification of the competitor onset for the first question and the amount of feature deviance for the second, plus the lexical statistics used as covariates for both. Lexical statistics were removed in a stepwise procedure if not significant.

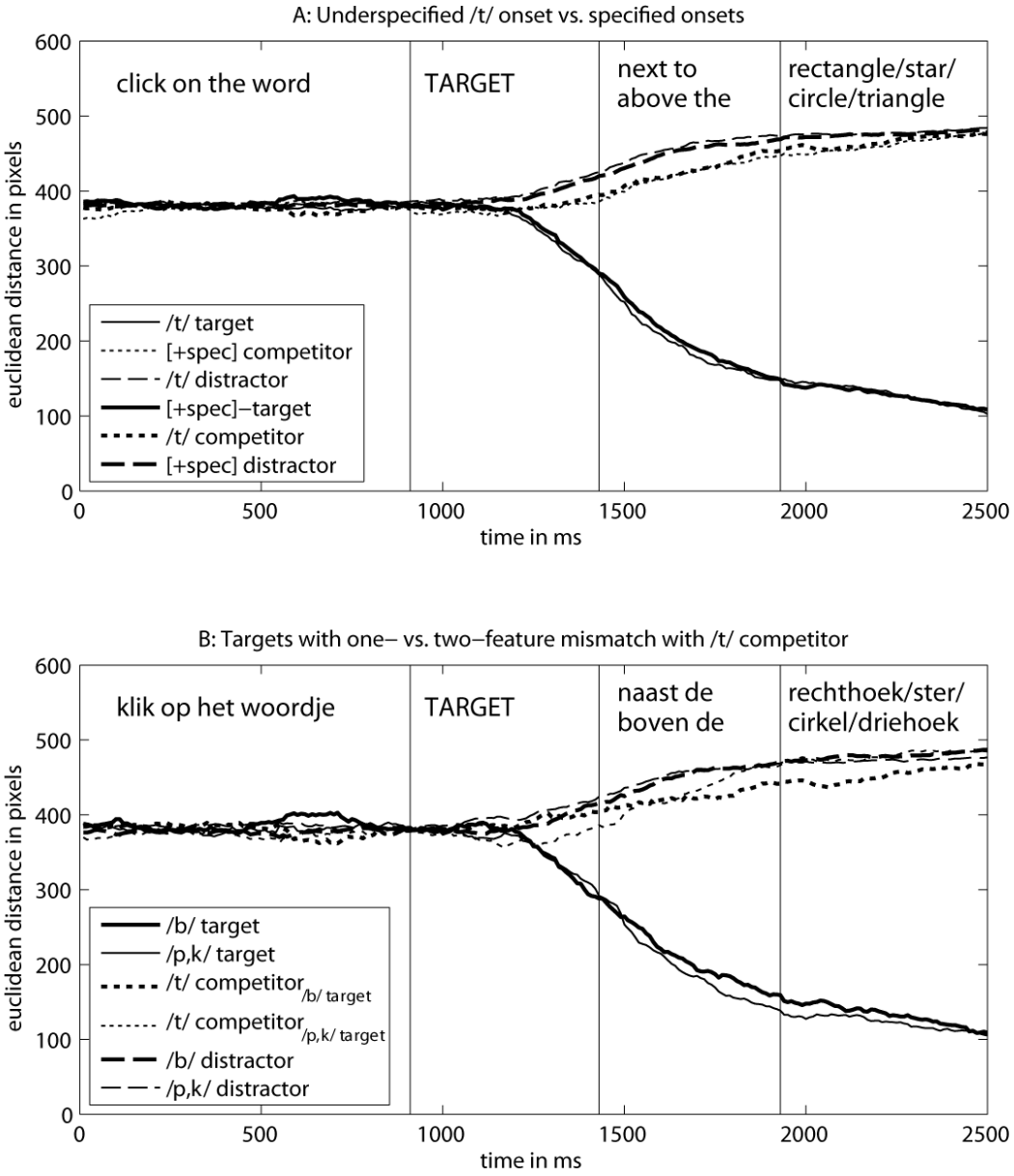

Figure 5. Eye-tracking results from the phonological trials in Experiment 4. The vertical lines in both panels indicate the onset of the different parts of the instructions. Panel A shows the comparison of looks to underspecified /t/ competitors and specified competitors. Panel B shows the comparison of looks to /t/ competitors that differ in one or two features from the target word.

Results

Performance was too close to ceiling to allow a meaningful error analysis (99.5% correct for trials with a /t/-initial target word, 100% for trials with another initial stop). Figure 5 shows the eye-tracking results for the correct trials, with the data for targets with specified vs. underspecified onsets in the

upper panel and the data for specified onsets split up in one- vs. two-feature mismatches in the lower panel. The upper panel displays a replication of earlier results. In the time window 200 to 600 ms after target onset, competitors received more looks than distractors, but there was no effect of the assumed specification of the onset consonant (B = -3.7, SE(B) = 7.4, t = -0.51, p_{MCMC} = 0.6). The lower panel, however, shows that words with an initial /t/ serving as competitors received more looks when they differed from the target in only one feature than when they differed in two features (B = 24 pixels, SE(B) = 10.9, t = 2.24, $p_{MCMC} < 0.05$, d = 0.15). In both analyses, lexical statistics had no significant influence on gaze behavior.

Discussion

This experiment revealed two data points that favor an optimal-perception account over the FUL model. The first result was a replication of the null result obtained in first three experiments. The activation of words with a less than perfect match with the input was independent of the assumed specification of the mismatch. Assumed "no-mismatches" (the word /tart/ with the input [part]) were not more strongly activated than assumed "mismatches" (the word /part/ with the input [tart]).

The second result that challenges the FUL model was that the number of mismatching default features influenced the activation of competitors. Competitors that only mismatch in voicing (e.g., /tart/ given the input [part]) seemed to be activated more strongly than competitors that mismatched in place and voicing (e.g., /tart/ given the input [bart]). This result may not be surprising, but it is not predicted by the FUL model. The lexical representation for the word /tart/ specifies neither a place nor voicing feature for the onset position. The only difference between the two situations is hence the number of "no mismatch" relations between input and lexical representation. The number of "no mismatch" relations, however, does not influence the lexical activation of word candidates (cf. formula (3) in Lahiri & Reetz, 2002).

It is also this result that clearly speaks for the assumption that spoken-word recognition makes use of all available information and hence behaves optimally. The reasonably small additional amount of feature mismatch between a one-feature mismatch and a two-feature mismatch is used to prevent strong activation of a spurious word candidate.

General Discussion

Four experiments pitted the predictions of the FUL model of spoken-word recognition against the predictions of an optimal-perception account. The FUL model predicts that the spoken-word recognition system uses underspecified lexical entries. As a consequence, not all of the information in the speech signal is used to the fullest extent possible. The theory assumes that the feature values [-Voice] for voicing and [Coronal] for place of articulation are default and hence not specified in the mental lexicon. This predicts an asymmetric matching between input and lexical representations. Inputs with non-default feature values do not mismatch lexical representations of words with

unspecified features in the same position. That is, the input [part] does not mismatch the lexical presentation of the Dutch word /tart/ (Engl. 'cake'). Asymmetric matching then arises, because inputs with default feature values mismatch the lexical representation of words with specified features in the same position. That is, the input [tart] mismatches the lexical representation of Dutch word /part/ (Engl. 'horse'). This was tested in four eye-tracking experiments with the visual-world paradigm. Participants heard an instruction to click on one of four printed words that were displayed on a computer screen in front of them. In contrast to the predictions of the FUL model, participants looked equally often at the printed word *taart* given the auditory input [part] as they looked at the printed word *paart* given the auditory input [tart]. This result was observed four times. Experiments 1 and 4 used minimal pairs (such as /part/-/tart/ 'horse' and 'cake'). While the mismatching words were more attractive than the distractor words, the effect was symmetric. Experiment 2 and 3 used words with a phonological overlap in the first syllable (such as /koxəl/-/toxa/, 'bullet' and 'toga'). In Experiment 2, both words appeared on the screen and one of them was the target. In Experiment 3, the printed word corresponding to the input was not in the screen in order to increase the amount of looks to the phonological competitor (Huettig & Altmann, 2005). Nevertheless, in both experiments, looks to the printed word *kogel* with the input [toxa] were as likely as looks to the printed word *toga* with the input [koxəl]. Conducting a meta-analysis over all experiments, the average *t* value for the predicted asymmetry is 0.20. This indicates that it is safe to assume that an asymmetric matching effect is absent.

Given that current statistical practice does not allow us to accept the null-hypothesis, a failure to find a statistically significant difference may be thought to indicate a failure of the research(er) rather than a failure of the theory (Kuhn, 1970). There are numerous problems with this argument. First of all, the standard error of the estimated (non-significant) effects were consistently small enough to lead to significant effects given the effect sizes that have been observed for phonological effects in eye-tracking experiments (Gow & McMurray, 2007; Mitterer & McQueen, 2009). Second, the eye-tracking methodology has proven quite sensitive and revealed phonological context effects that were not apparent in behavioral data (Mitterer & Ernestus, 2006; Mitterer & McQueen, 2009). Third, the current enterprise provided not one but four tests of the asymmetric-matching hypothesis. None of them revealed a significant effect, and the overall effect is close to zero.

Nevertheless, Experiment 4 "reversed" the roles of the FUL model and an optimal-perception account, so that the FUL-model predicts no effect but an optimal-perception account does. An optimal perception account predicts that listeners are sensitive to small phonetic differences, so that the estimated posterior probability of a target word is smaller if the input mismatches in two features than when the input mismatches in only one feature. The FUL model predicts such an effect only if the features are specified, while the number of "non-mismatching" features is irrelevant for lexical access. The Dutch word /tart/ ('cake') is supposed to be activated to the same degree by the input [part]—

(non)-mismatching in place—as by the input [bart]—(non)-mismatching in place and voicing. The results followed the predictions of an optimal-perception account and showed that the word /tart/ is more strongly activated by the input [part] than by the input [bart].

The question then arises whether an optimal-perception account can account for other patterns of data taken as evidence for underspecification. Most prominently, Lahiri and Marslen-Wilson (1991) investigated how Bengali and English listeners interpret nasalized vowels in a gating task. Importantly, Bengali listeners can interpret a nasalized vowel as an inherently nasal vowel in a CṼC word or as an underlyingly oral vowel in a CVN word, which is produced as CṼN. The results indicated that Bengali listeners interpreted the nasalized vowel mostly as a nasal vowel in a CṼC word. Lahiri and Marslen-Wilson took this as evidence that nasalization is specified and oral vowels are underspecified. As argued elsewhere (McQueen, 1995; Ohala & Ohala, 1995), these results are, however, less clear cut than underspecification theory assumes. Moreover, the preference for CṼC words given a nasal vowel may alternatively show that perceivers tend to have a preference for more tangible causes. That is, when hearing a nasal vowel, it is a safer bet to assume that this is because the vowel already heard is nasal than assuming that the upcoming consonant is nasal. That is, listeners prefer context-independent judgments. This is similar to a set of findings in color perception, in which perceivers tend to assume that an orange plate is indeed orange. Only when the possibility of a context, such as a red illuminant, is highlighted, are perceivers more likely to judge the plate as yellow (Arend & Reeves, 1986; Troost & de Weert, 1991).

An optimal-perception account can hence deal with some of the evidence for the underspecification account, but clearly not all. Those datasets that support an asymmetric matching (Eulitz & Lahiri, 2004; Friedrich, et al., 2006, 2008; Lahiri & Reetz, 2002) clearly contradict the assumption of an optimal-perception account. In such a situation, it is important to note that negative evidence weighs heavier than positive evidence in scientific research. To use a common example, a theory that "all swans are white" is not confirmed by observations of white swans. In the current context, the reported asymmetries are analogous to such observations of white swans. Given the now massive negative evidence—observed symmetries where the FUL model predicts asymmetries—obtained with priming measures (Gow, 2001), EEG and MEG recordings (Bonte et al., Tavabi et al. 2009), and the current evidence from eye-tracking, it seems that the theory of phonological underspecification cannot be maintained in its current form. Moreover, the existence of asymmetries is widespread (Polka & Bohn, 2003) and may be explained by other devices. Darcy and Kügler (2007), for instance, found more compensation for voicing assimilation than for devoicing assimilation in French listeners. Instead of appealing to underspecification, they explained the data pattern on the basis of the production pattern in French.

Nevertheless, one of the foundations of the FUL model remains a valid point: Coronal consonants seem to constitute a special case. They are the most likely to undergo changes in connected speech, such as flapping in American English (Tucker & Warner, 2007). Moreover, alveolar consonants are also extremely "popular" in the phonological inventories of the languages of the world (Paradis & Prunet, 1991), although not universal (Blevins, 2009) as assumed by underspecification theory (Lahiri & Reetz, 2010). In a similar vein, the current experiments showed that phonological underspecification is probably not the best way to conceptualize this special status. In fact, by remaining focused on the question of whether coronal segments are specified in the mental lexicon or not, it is difficult to consider new ideas of how the special status of coronals could be better captured.

Another strong point in the original conception of underspecification is that it allows listeners to recognize reduced variants. However, the FUL model does not capture the context effects in the perception of assimilated forms. Other recent work on the recognition of variants is also problematic for the assumptions of the FUL model. Work on medial schwa deletion ("camra" for *camera*) indicates that variant frequency may be encoded in the mental lexicon (Connine, Ranbom, & Patterson, 2008), contrasting with purely phonological and abstract lexical entries in the FUL model. In a similar vein, Pitt (2009) showed that exposure to flapped variants ("winner" for *winter*) is essential for word recognition. Additionally, Sumner and Samuel (2005) found a "canonical form advantage" for variants of word-final /t/. That is, recognition was most efficient if the word was uttered in its canonical form. This finding contrasts with assumption that the presence or absence of default features in the input does not influence the activation level of words with underspecified segments. Finally, context effects in the perception of /t/-reduction are also difficult to explain in the framework of the FUL model (Mitterer & Ernestus, 2006; Mitterer & McQueen, 2009).

In summary, the current paper provides a large set of problematic empirical data for the FUL model. One of the original reasons to assume underspecification was that it helps the listener to deal with the recognition of phonological variants. However, the model turns out to be incompatible with a growing body of research on the perception of phonological variants. Empirically and conceptually then, there is little reason to maintain the assumption of phonological underspecification in the mental lexicon.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439.

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology, 6*, 183-207.

Arend, L., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America, 3*, 1743-1751.

Baayen, H. R. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge, UK: Cambridge University Press.

Blevins, J. (2009). Another universal bites the dust: Northwest Mekeo lacks coronal phonemes. *Oceanic Linguistics, 48*, 264-273.

Bonte, M., Mitterer, H., Zellagui, N., Poelmans, N., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology, 116*, 2765.

Cairns, C. E. (1988). Phonotactics, markedness and lexical representation. *Phonology, 6*, 209-236.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335-359.

Coenen, E., Zwitserlood, P., & Bölte, J. (2001). Consequences of assimilation for word recognition and lexical representation. *Language and Cognitive Processes, 15*, 535 - 564.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition. *Journal of Memory and Language, 32*, 193-210.

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics, 70*, 403-411.

Cutler, A., & Broersma, M. (2005). Phonetic precision in Listening. In W. J. Hardcastle & J. Mackenzie Beck (Eds.), *A figure of speech: a Festschrift for John Laver* (pp. 63-91). London: Lawrence Earlbaum.

Darcy, I., & Kügler, F. (2007). Phonological context effects for voicing and devoicing in French. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences.* (pp. 1257-1260). Saarbrücken: Diderot.

Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules: perceptual compensation for assimilation in late L2-learners. In J. Cole & J. Hualde (Eds.), *Laboratory Phonology 9* (pp. 411-442). Berlin: Mouton de Gruyter.

Deacon, D., Hewitt, S., Yang, C.-M., & Masanouri, N. (2000). Event-related potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical. *Cognitive Brain Research, 9*, 137-146.

Ernestus, M., & Mak, W. M. (2004). Distinctive phonological features differ in relevance for both spoken and written word recognition. *Brain and Language, 90*, 378-392.

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience, 16*, 577-583.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America, 99*, 1730-1741.

Friedrich, C. K., Lahiri, A., & Eulitz, C. (2006). Not every pseudoword disrupts word recognition: An ERP study. *Behavioral and Brain Functions, 2*, 26.

Friedrich, C. K., Lahiri, A., & Eulitz, C. (2008). Neurophysiological evidence for underspecified lexical representations: Asymmetries with word initial variations. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1545-1559.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 144-158.

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 380-396.

Gaskell, M. G., & Snoeren, N. D. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1632-1647.

Geisler, W. S., & Kersten, D. (2002). Illusions, perception and Bayes. *Nature Neuroscience, 5*, 508-510.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45*, 133-159.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 163-179.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics, 65*, 575-590.

Gow, D. W., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. In J. Cole & J. Hualde (Eds.), *Laboratory Phonology 9* (pp. 173-200). New York: Mouton de Gruyter.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance (Ch. 10)* (pp. 1-112). New York: Wiley.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition, 96*, B23-B32.

Jacobsen, T., Schröger, E., Winkler, I., & Horvath, J. (2005). Familiarity affects thte processing of task-irrelevant auditory deviants. *Journal of Cognitive Neuroscience, 17*, 1704-1713.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium* (pp. 29-54). Tokyo, Japan: The National International Institute for Japanese Language.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions, 2nd. ed.* Chicago: University of Chicago Press.

Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the lexicon. *Cognition, 38*, 245-294.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 637-676). Berlin: Mouton de Gruyter.

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing *Journal of Phonetics, 38*(1), 44-59.

Marslen- Wilson, W. D., Nix, A., & Gaskell, M. G. (1995). Phonological variation in lexical access: abstractness, inference and English place assimilation. *Language and Cognitive Processes, 10*, 285-308.

McQueen, J. M. (1995). Processing versus representation: comments on Ohala and Ohala. In B. Cornell & A. Arvanti (Eds.), *Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV* (pp. 61-67). Cambridge: Cambridge University Press.

McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology, 60*, 661-671.

Mitterer, H. (2003). *Understanding "gardem bench": Studies on the perception of assimilation word forms [dissertation]*. Maastricht, The Netherlands: Universiteit Maastricht.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics, 65*(6), 956-969.

Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the perception of assimilated word forms. *Quarterly Journal of Experimental Psychology, 59*, 1395-1424.

Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of assimilated word forms does not depend on specific language experience. *Cognitive Science, 30*, 451-479.

Mitterer, H., & de Ruiter, J. P. (2008). Recalibrating color categories using world knowledge. *Psychological Science, 19*, 629-634.

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., & 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.

Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.

Näätänen, R. (1995). The Mismatch Negativity: A powerful tool for cognitive neuroscience. *Ear & Hearing, 16*, 6-18.

Näätänen, R., Pakarinena, S., Rinnea, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clinical Neurophysiology, 115*(140-14).

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neuroscience, 24*(283-289).

Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. In G. Docherty & D. R. Ladd (Eds.), *Laboratory Phonology 2* (pp. 261-280). Cambridge: Cambridge University Press.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian Model of Continuous Speech Recognition. *Psychological Review, 115*, 357-395.

Ohala, J. J., & Ohala, M. (1995). Speech perception and lexical representation: The role of vowel nasalization in Hindi and English. In B. Cornell & A. Arvanti (Eds.), *Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV* (pp. 41-60). Cambridge: Cambridge University Press.

Paradis, C., & Prunet, J.-F. (1991). Introduction: Asymmetry and visibility in consonant articulations. In C. Paradis & J.-F. Prunet (Eds.), *Phonetics and Phonology: The Special Status of Coronals Internal and External Evidence (, Vol. 2)* (2 ed., pp. 1-28). San Diego: Academic Press.

Pitt, M. A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language, 61*(19-36).

Pitt, M. A., & Samuel, A. G. (2006). Word length and lexical activation: longer is better. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1120-1135.

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication, 41*, 221-231.

Repp, B. H., & Svastikula, K. (1988). The perception of the [m]-[n] distinction in VC syllables. *Journal of the Acoustical Society of America, 83*, 237-247.

Scharinger, M. (2009). Minimal representations of alternating vowels. *Lingua, 119*, 1414-1425.

Schröger, E., Bendixen, A., Trujillo-Barreto, N. J., & Roeber, U. (2007). Processing of abstract rule violations in audition. *Plos ONE, 2, e1131., 2*, e1131.

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.

Tavabi, K., Elling, L., Dobel, C., Pantev, C., & Zwitserlood, P. (2009). Effects of place of articulation changes on auditory neural activity: a magnetoencephalography study. *Plos ONE, 4*(2), e4452.

Troost, J. M., & de Weert, C., M.M. (1991). Naming vs. matching in color constancy. *Perception & Psychophysics, 50*, 591-602.

Tucker, B. V., & Warner, N. (2007). Inhibition of processing due to reduction of the American English flap. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken: Pirot.

van der Feest, S. V. H. (2007). *Building a phonological lexicon [dissertation]*. Nijmegen: Radboud Universiteit.

Van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*, 241-255.

Wheeldon, L., & Waksler, R. (2004). Phonological underspecification and mapping mechanisms in the speech recognition lexicon. *Brain and Language, 90*, 401-412.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences, 10*, 301-308.

Footnotes

[1] I will use the term "underspecified" for words or phonemes for which only some features are specified and "unspecified" for features that are not specified.

[2] It is important to note that I do not intend to pit the FUL model against the Shortlist B model of word recognition. The assumption of "optimal perception"—as contrasted with the assumption of phonological underspecification—is in fact implicitly present in most models of speech perception and spoken-word recognition. Consider the direct-realist theory of speech perception (e.g., Fowler, 1996). On a meta-theoretical level, this theory is strongly opposed to assumption of perceptual inference that lies at the heart of Bayesian models of perception. Nevertheless, for the current purposes, it still falls under the umbrella of models that assume "optimal perception", because it assumes that listeners will make use of all information in the speech signal to the fullest extent.

[3] There is no consensus yet how to define effect size measures in linear mixed-effect models. Hence, I report the ratio of the estimated unstandardized regression coefficient and the standard deviation of the dependent variable as Cohen's *d* for categorical predictors. For continuous predictors, such as lexical frequency, I present the amount of explained variance on the basis of the correlation between a predictor and the residuals of the model without that predictor. This measure is similar to the squared partial eta.

Appendix

Table A1: Materials for the phonological trials Experiment 1.

Specified initial segment	Coronal initial segment	IPA
fijne	seinen	/{f,s}ɛinə/
fabel	sabel	/{f,s}abəl/
foppen	soppen	/{f,s}ɔpə/
fik	sik	/{f,s}ɪk/
fauna	sauna	/{f,s}ɑuna/
fel	cel	/{f,s}ɛl
fier	sier	/{f,s}ir/
fijn	sein	/{f,s}ɛin/
flipper	slipper	/{f,s}lɪpər/
flank	slank	/{f,s}lɑŋk/
flap	slap	/{f,s}lɑp/
fleuren	sleuren	/{f,s}lørə/
flik	slik	/{f,s}lɪk/
flikken	slikken	/{f,s}lɪkə/
flippen	slippen	/{f,s}lɪpə/
floep	sloep	/{f,s}lup/
flop	slop	/{f,s}lɔp/
fluiten	sluiten	/{f,s}lœytə/
fluiter	sluiter	/{f,s}lœytər/
fok	sok	/{f,s}ɔk/
mul	nul	/{m,n}ʏl/
muffig	nuffig	/{m,n}ʏfəx/
mouw	nauw	/{m,n}ɑuw/
motie	notie	/{m,n}otsi/
mop	nop	/{m,n}ɔp/
moorden	noorden	/{m,n}ordə/
Moor	Noor	/{m,n}or/
mol	nol	/{m,n}ɔl/
mok	nok	/{m,n}ɔk/
mode	node	/{m,n}odə/
mixen	niksen	/{m,n}ɪksə/
mis	nis	/{m,n}ɪs/
meer	neer	/{m,n}er/
mier	nier	/{m,n}ir/
mest	nest	/{m,n}ɛst/
mep	nep	/{m,n}ɛp/
meid	nijd	/{m,n}ɛit/
mat	nat	/{m,n}ɑt/
macht	nacht	/{m,n}ɑxt/
maat	naad	/{m,n}at/
bank	dank	/{b,d}ɑŋk/
bal	dal	/{b,d}ɑl
boos	doos	/{b,d}os
boot	dood	/{b,d}ot/
buur	duur	/{b,d}yr/
bom	dom	/{b,d}ɔm/
bak	dak	/{b,d}ɑk/
bol	dol	/{b,d}ɔl/
borst	dorst	/{b,d}ɔrst/

bon	don	/{b,d}ɔn/
boel	doel	/{b,d}ul/
baat	daad	/{b,d}at/
bier	dier	/{b,d}ir/
bekken	dekken	/{b,d}ɛkə/
beuk	deuk	/{b,d}øk/
boek	doek	/{b,d}uk/
bot	dot	/{b,d}ɔt/
bril	dril	/{b,d}rɪl/
braak	draak	/{b,d}rak/
broom	droom	/{b,d}rom/

Table A2. Materials used in Experiments 2 and 3.

Specified first segment	IPA	Coronal first segment	IPA	Overlap
baden	badə	dadel	dadəl	4
baklucht	bɑkluxt	dakloos	daklos	4
balie	bali	daling	dalɪŋ	4
bamboe	bɑmbu	dambord	dɑmbɔrt	4
beginnen	bexɪnə	degene	dəxene	3
beker	bekər	deken	dekə	3
bekken	bɛkə	dekking	dɛkɪŋ	3
bemanning	bəmɑnɪŋ	dementie	dəmentsi	3
beurs	børs	deuk	døk	2
biefstuk	bifrstʏk	diefstal	difstɑl	4
bierbuik	birbœyœyk	dierbaar	dirbar	4
bijbel	bɛibəl	dijbeen	dɛiben	3
bizon	bisɔn	diesel	disəl	3
bode	bodə	doding	dodɪŋ	3
boeking	bukɪŋ	doeken	dukə	3
boeman	bumɑn	doemen	dumə	3
bokkig	bɔkəx	dokken	dokə	4
bolling	bɔliŋ	dollen	bɔlə	3
fabricage	fabrikaʃe	sabotage	sabotaʃe	3
failliet	fɑjɪt	saillant	sɑjɑnt	3
fakkeltocht	fɑkəltɔxt	sakkerloot	sɑkərlot	4
falen	falə	salie	sali	3
familie	famili	salami	salami	2
fanatiekeling	fanatikəlɪŋ	sanatorium	sanatorijʏm	4
fatalisme	fatɑlɪsmə	satelliet	satəlit	3
fatsoen	fatsun	satijn	satɛin	3
federaal	federal	sediment	sedimɛnt	3
felomstreden	fɛlɔmstredə	selderijsoep	sɛldərɛisup	3
fervent	fɛrfɛnt	serveer	server	3
fierheden	firhedə	sierheester	sirhestər	4
figuurlijk	fixyrlək	sigaren	sixaren	3
fijnheid	fɛinhɛit	seinhoorn	sɛinhorn	5
filiaal	filijal	silhouet	siluwɛt	3
fiscus	fɪskʏs	sisklank	sisklank	4
fluiten	flœytə	sluiting	slœytɪŋ	4
kaakstoot	kakstot	taakstraf	takstrɑf	5
kaalte	kaltə	taaltoets	taltuts	4
kapping	kɑpɪŋ	tappen	tɑpə	3
koers	kurs	toef	tuf	2
kogel	koxəl	toga	toxa	3
koning	konɪŋ	tonen	tonə	3
maan	man	naald	nalt	2
mager	maxər	nagels	naxəls	3
meegaand	mexant	negen	nexə	3
meeuw	mew	neef	nef	2
mergel	merxəl	nergens	nɛrxəns	5
mogen	moxə	noga	noxa	3
praktisch	prɑktɪs	tractor	trɑktɔr	5

Table A3. Items with 1 vs. 2 feature difference to a /t/-inital word.

1 feature difference	2 features difference	/t/-initial word	IPA
cross	bros	tros	/{k,b,t}ɔs/
kaal	baal	taal	/{k,b,t}al/
kaken	baken	taken	/{k,b,t}akə/
kalk	balk	talk	/{k,b,t}ɑlk/
kauwen	bouwen	touwen	/{k,b,t}ɑuwə/
keer	beer	teer	/{k,b,t}er/
kei	bij	tij	/{k,b,t}ɛi/
keken	beken	teken	/{k,b,t}ekə/
kent	band	tent	/{k,b,t}ɛnt/
kermen	bermen	termen	/{k,b,t}ɛrmə/
kieren	bieren	tieren	/{k,b,t}irə/
kikken	bikken	tikken	/{k,b,t}ɪkə/
kil	bil	til	/{k,b,t}ɪl/
kind	bind	tint	/{k,b,t}ɪnt/
kocht	bocht	tocht	/{k,b,t}ɔxt/
koe	boe	toe	/{k,b,t}u/
kok	bok	tok	/{k,b,t}ɔk/
komen	bomen	tomen	/{k,b,t}omə/
kon	bon	ton	/{k,b,t}ɔn/
koren	boren	toren	/{k,b,t}orə/
kracht	bracht	tracht	/{k,b,t}rɑxt/
krant	brand	trant	/{k,b,t}rɑnt/
krauwen	brouwen	trouwen	/{k,b,t}ɑuwə/
kreet	breed	treedt	/{k,b,t}ret/
krijn	brein	trein	/{k,b,t}rɛin/
krom	brom	trom	/{k,b,t}rɔm/
krui	brui	trui	/{k,b,t}rœy/
kuil	buil	tuil	/{k,b,t}œyl/
kuis	buis	thuis	/{k,b,t}œys/
kuit	buit	tuit	/{k,b,t}œyt/
paai	baai	taai	/{p,b,t}ai/
paard	baard	taart	/{p,b,t}art/
pakken	bakken	takken	/{p,b,t}ə/
pal	bal	tal	/{p,b,t}ɑl/
pand	band	tand	/{p,b,t}ɑnt/
pas	bas	tas	/{p,b,t}ɑs/
pellen	bellen	tellen	/{p,b,t}elə/
penen	benen	tenen	/{p,b,t}enə/
pips	bips	tips	/{p,b,t}ɪps/
poef	boef	toef	/{p,b,t}uf/
poen	boen	toen	/{p,b,t}on/
pol	bol	tol	/{p,b,t}ɔl/
poon	boon	toon	/{p,b,t}on/
pop	bob	top	/{p,b,t}ɔp/
pot	bot	tot	/{p,b,t}ɔt/
pril	bril	tril	/{p,b,t}rɪl/
pui	bui	tui	/{p,b,t}œy/