
ELSEVIER

Speech Communication 41 (2003) 257–270

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

# Flow of information in the spoken word recognition system ☆

James M. McQueen [a,*], Anne Cutler [a], Dennis Norris [b]

[a] *Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands*
[b] *MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 2EF, UK*

**Abstract**

Spoken word recognition consists of two major component processes. First, at the prelexical stage, an abstract description of the utterance is generated from the information in the speech signal. Second, at the lexical stage, this description is used to activate all the words stored in the mental lexicon which match the input. These multiple candidate words then compete with each other. We review evidence which suggests that positive (match) and negative (mismatch) information of both a segmental and a suprasegmental nature is used to constrain this activation and competition process. We then ask whether, in addition to the necessary influence of the prelexical stage on the lexical stage, there is also feedback from the lexicon to the prelexical level. In two phonetic categorization experiments, Dutch listeners were asked to label both syllable-initial and syllable-final ambiguous fricatives (e.g., sounds ranging from [f] to [s]) in the word–nonword series *maf–mas*, and the nonword–word series *jaf–jas*. They tended to label the sounds in a lexically consistent manner (i.e., consistent with the word endpoints of the series). These lexical effects became smaller in listeners' slower responses, even when the listeners were put under pressure to respond as fast as possible. Our results challenge models of spoken word recognition in which feedback modulates the prelexical analysis of the component sounds of a word whenever that word is heard.


**Zusammenfassung**

Das Erkennen gesprochener Wörter besteht aus zwei Hauptverarbeitungskomponenten. Auf der prälexikalischen Ebene wird die Information im Sprachsignal genutzt, um eine abstrakte Beschreibung der Äußerung zu generieren, welche dann verwendet wird, um auf gespeicherte lexikalische Information zuzugreifen. Die lexikalische Ebene ist zum einen gekennzeichnet durch die vielfältige Aktivierung von Wort-Kandidaten, die mit dem Sprachinput übereinstimmen, und zum anderen durch den 'Wettbewerb' dieser Kandidaten untereinander. Wir werden zunächst Belege rezensieren, die nahe legen, dass sowohl positive (übereinstimmende) als auch negative (nicht übereinstimmende) Information—sowohl segmentaler wie suprasegmentaler Natur—genutzt wird, um den Aktivierungs- und Wettbewerbsprozess zu steuern. Danach beschäftigen wir uns mit der Frage, ob zusätzlich zum notwendigen Einfluss der prälexikalischen auf die lexikalische Ebene, ebenfalls Feedback vom Lexikon zum prälexikalischen Level stattfindet. In zwei phonetischen Kategorisierungsexperimenten sollten niederländische Hörer silben-initiale und silben-finale ambige Frikative (z.B., Laute zwischen [f] und [s]) in der Wort–Nichtwort Reihe *maf–mas* und der Nichtwort–Wort Reihe

☆ A preliminary report of these experiments was presented at the 137th meeting of the Acoustical Society of America in Berlin, Germany, in March 1999.

* Corresponding author. Tel.: +31-24-3521373; fax: +31-24-3521213.
*E-mail addresses:* james.mcqueen@mpi.nl (J.M. McQueen), anne.cutler@mpi.nl (A. Cutler), dennis.norris@mrc-cbu.cam.ac.uk (D. Norris).

0167-6393/02/$ - see front matter 
doi:10.1016/S0167-6393(02)00108-5

*jaf–jas* benennen. Sie neigten dazu, diese Laute so zu benennen dass sie lexikalisch konsistent waren (d.h., konsistent mit dem jeweiligen Wort-Endpunkt der beiden Kontinua). Dieser lexikalische Effekt wurde in den langsameren Reaktionszeiten geringer, auch wenn die Hörer unter Druck gesetzt wurden, so schnell wie möglich zu antworten. Unsere Ergebnisse stellen Sprachverarbeitungsmodelle in Frage, in denen Feedback bei jedem Hören eines Wortes die prälexikalische Analyse einzelner Laute dieses Wortes beeinflußt.


**Résumé**

La reconnaissance d'un mot parlé consiste en deux étapes principales. Lors de l'étape pré-lexicale, l'information présente dans le signal de parole est analysée, et la représentation abstraite de l'énoncé ainsi générée est utilisée lors de l'accès au lexique mental. L'étape lexicale se caractérise par l'activation simultanée des candidats lexicaux qui sont consistents avec l'information sensorielle, et ces candidats rentrent en compétition. Cet article présente des données indiquant que les informations segmentales et suprasegmentales présentes dans le signal de parole peuvent jouer un rôle positif ou négatif lors de ce processus de compétition selon que ces informations sont compatibles ou incompatibles avec la représentation lexicale des candidats. De plus, nous discutons l'existence d'un "feedback" du lexique vers le niveau pré-lexicale qui s'ajouterait ainsi au flux d'information nécessaire entre le niveau pré-lexicale et le lexique. Lors de deux experiences de catégorisation phonétique, conduites en néerlandais, les auditeurs avaient pour tâche de décider de l'identité d'une fricative ambiguë présente au début ou à la fin d'une séquence qui correspondait, selon l'identité attribuée à cette fricative, soit à un mot, soit ou un nonmot (e.g., une fricative ambiguë entre [f] et [s] placée dans une séquence mot–nonmot, *maf–mas*, ou dans une séquence nonmot–mot, *jaf–jas*). Un biais lexical était observé. Ce biais lexical était, cependant, plus faible lorsque les temps de réaction lors de la catégorisation étaient lents, même lorsque les auditeurs étaient encouragés à répondre le plus vite possible. Ces résultats sont problématiques pour les modèles dans lesquels l'analyse pré-lexicale du signal sonore est modulée par un "feedback" opérant systématiquement, toutes les fois où un mot est perçu.


*Keywords:* Spoken word recognition; Levels of processing; Feedback; Phonetic categorization

## 1. Introduction

Most models of spoken word recognition, including TRACE (McClelland and Elman, 1986), Shortlist (Norris, 1994), the distributed cohort model (DCM; Gaskell and Marslen-Wilson, 1997) and PARSYN (Luce et al., 2000), make the assumption that word recognition is fundamentally a two-stage process. First, there is a prelexical stage, where an abstract description of a given utterance is generated. There is no agreement about the form of this description (features and then, at a second stage, phonemes in TRACE; phonemes in Shortlist; features in the DCM; allophones in PARSYN), but broad agreement that there are processes of abstraction and normalization which mediate between the speech signal and the mental lexicon. The second, lexical component of the recognition process involves the activation of many candidate words (those which match the prelexical representation of the input to some extent), and competition among those words. The words which win the competition are recognized. Again, while there are many disagreements about the details of the activation–competition process, TRACE, Shortlist, the DCM and PARSYN all assume activation and competition in some form.

In the first part of this paper we discuss recent results which bear on two disputed aspects of the activation–competition process, namely, whether both positive and negative information can be used to constrain the set of candidate words, and whether the competition process entails direct inhibition between candidate words. These results also put important constraints on the kind of information that is represented at both the prelexical and lexical levels of processing. In the second part of the paper we discuss whether there is feedback from the lexical to the prelexical level.

In Shortlist, but not TRACE for example, information at the prelexical level influences lexical activation both positively (a match between the information in the signal and stored phonological knowledge facilitates lexical activation) and negatively (mismatch between the signal and the lexicon inhibits activation). Furthermore, in Shortlist, but not in the DCM, for example, there is direct competition between candidate words. There are therefore two mechanisms in Shortlist by which erroneous candidates can be rejected: bottom-up inhibition and lexical competition. While there are many demonstrations that multiple candidates which match the signal either temporarily or partially are activated, and that they compete with each other (see Frauenfelder and Floccia, 1998, for a review), it is not yet clear what the relative contributions of bottom-up inhibition and lexical competition are to the resolution of the recognition process. Recent results suggest that bottom-up mismatch can be strongly inhibitory (Frauenfelder et al., 2001); other results suggest that lexical competition resolves very soon after disambiguating information becomes available (McQueen et al., 1999).

Soto-Faraco et al. (2001) have recently examined this issue. They also considered an additional variable: the relative roles of segmental and suprasegmental information in lexical access (i.e., the relative contributions to a word's activation of its component phonemes and of its lexical stress pattern). Spanish listeners heard Spanish sentences which ended with word fragments, and then made lexical decisions on visual target words (and nonwords) which were presented at the offset of the fragments. When the target words matched the fragments, responses were facilitated (relative to when they followed phonologically unrelated control fragments). When the target words mismatched the fragments, responses were inhibited relative to control. The amount of inhibition was equivalent across a range of different mismatch conditions: lexical stress mismatch (e.g., *prinCI-*, the beginning of *prinCIpio*, followed by the target PRINCIPE, which is stressed on the first syllable, *PRINcipe*); vowel mismatch (e.g., *abun-*, from *abundancia*, followed by ABANDONO); and consonant mismatch (e.g., *pati-*, from *patilla*, followed by PAPILLA).

These inhibition effects suggest the following: both of the words are activated by the spoken input (e.g., *abundancia* and *abandono* given *abun-*), the matching word is favored and the mismatching word is disfavored (by bottom-up inhibition), and lexical competition further penalizes the mismatching candidate, producing relatively slow responses to visual versions of that mismatching word. Although inhibition can of course arise by bottom-up mismatch alone, the added contribution of competition between words to the inhibition observed by Soto-Faraco et al. (2001) can be seen by comparing their results with those of Cutler and Donselaar (2001), who conducted a similar fragment-priming study in Dutch. In that study, no inhibition effects were observed. Listeners were faster to decide, for example, that MUSEUM was a word after hearing the matching fragment *muZEE-* than after hearing a control fragment. They were also faster after the fragment *MUzee* (stress mismatch) than in the control condition, though the facilitation here was weaker than in the matching condition; crucially, however, there was no difference between control and segmental mismatch conditions (e.g., *luZEE-*).

While the difference between the segmental and suprasegmental mismatch conditions in Dutch contrasts with equivalent effects for these conditions in Spanish, and may reflect the relative role of suprasegmental information in lexical access in these two languages (see Cutler and Donselaar, 2001; Soto-Faraco et al., 2001, for discussion), the most important difference between the two studies is that inhibition was observed in the mismatch conditions in Spanish but not in Dutch. The reason for this difference is that there were rival words activated by the stimuli in the Spanish study but not in the Dutch study (*prinCI-* is the beginning of *prinCIpio*, and *abun-* is the beginning of *abundancia*, but neither *MUzee-* nor *luZEE-* is the beginning of any Dutch word). The inhibition in Spanish therefore appears to reflect the added contribution of lexical competition between the representation of the rival word (consistent with the spoken input in the mismatch condition) and the representation of the target word.

Taken together, the results of Soto-Faraco et al. (2001) and Cutler and Donselaar (2001) suggest

that both bottom-up and lexical competition are involved in the word recognition process. They therefore support the assumptions about information flow built into the Shortlist model. They also suggest, however, that Shortlist (and indeed all other current models of spoken word recognition) should be revised such that suprasegmental information can be represented at both the lexical and the prelexical levels, at least where it is of value in constraining lexical access in a particular language. Note that studies in English, such as that of Cutler (1986), have so far failed to reveal suprasegmental effects on lexical access; this may be because suprasegmental information has low value in constraining lexical access in English (Cutler et al., 1997).

In addition to these issues about the bottom-up and lateral flow of information in the spoken word recognition system, a further important question needs to be answered: Does information also flow top-down? That is, is there feedback from the lexical level back down to the prelexical level of analysis? This question has exercised psycholinguists for more than two decades. There have been many demonstrations of lexical effects in tasks which require listeners to make explicit decisions about speech sounds. Thus, in phoneme monitoring, listeners can be faster to decide that they have heard a target sound, such as /b/, in a real word like *bid* than in a nonword like *bip* (Cutler et al., 1987; Rubin et al., 1976). Lexical effects have also been observed in the phoneme restoration illusion. Listeners are worse at determining whether a phoneme has been replaced by noise or had noise added to it in a real word than in a nonword (Samuel, 1981, 1996). One way to explain such effects is to assume that decisions about speech sounds are made at the prelexical level, and thus that the effects reflect feedback of lexical information to that level. This assumption is built into the TRACE model, where word representations are activated bottom-up when their constituent phonemes become activated, and in which word activation in turn modulates the activation of those phoneme representations via top-down connections.

Lexical effects can however be explained without postulating feedback. In the Merge model (McQueen et al., 1999; Norris et al., 2000), activation of representations at both the prelexical and lexical levels is continuously integrated at dedicated phonemic decision units. Merge shares with Shortlist all architectural assumptions about the word recognition process, but it is a model of phonemic decision-making rather than of word recognition (Norris et al., 2000). Lexical effects in words arise in Merge because activation at the lexical level increases the activation of the constituent phonemes of those words at the decision stage. Lexical effects in nonwords are rarer, but are nonetheless now well documented (Connine et al., 1997; Marslen-Wilson and Warren, 1994; McQueen et al., 1999; Newman et al., 1997). These effects arise because words which sound similar to those nonwords will be activated and again can bias the activation of the decision units. The simple demonstration of lexical involvement in phonemic decision-making therefore cannot be taken as prima facie support for feedback.

It is clear that an answer to the question about feedback will only be forthcoming if more detailed predictions of the available models are tested. An important first step, however, is to distinguish between different forms of feedback. Psycholinguists have been almost exclusively concerned with the question of whether feedback from a specific lexical item can affect the perceptual analysis of the prelexical units currently providing the input to that word. The effects of this feedback are transitory, and have no consequences for the later processing of that word or any other word. This is the nature of the feedback embodied in TRACE; we will refer to it as *perceptual feedback*. Another form of feedback, which we will refer to as *attentional feedback*, is generalised top-down attentional control over prelexical processing. For example, listeners could devote more or less attention to prelexical processing. Proponents of modular, bottom-up theories of perception (e.g., Norris et al., 2000; Pylyshyn, 1999) readily acknowledge that attentional factors can modulate the operation of early perceptual processes. There is no disagreement about the existence of this particular type of feedback. There is, however, a third form of feedback that has received little discussion in the word recognition literature: *feed-*

*back for learning*. Lexical knowledge could influence the way in which we learn perceptual categories. That is, lexical feedback need not have the immediate and specific effect on prelexical processing it does in TRACE, but might have the longer-term and more general effect of retuning prelexical processing. We will return to the issue of perceptual learning later in the paper. In the experiments reported here, we examined the predictions that follow from the perceptual feedback in TRACE. The way feedback is implemented in TRACE makes predictions about the time-course of lexical involvement in phonetic categorization.

Lexical information influences the categorization of ambiguous speech sounds in syllable-initial position (e.g., stops between [t] and [d] in the word–nonword (W–NW) series *type–dype* tend to be heard as [t]; Ganong, 1980; Connine and Clifton, 1987) and in syllable-final position (e.g., fricatives between [s] and [ʃ] in *kiss–kish* tend to be heard as [s]; McQueen, 1991). In the categorization of syllable-initial sounds, lexical effects get stronger in slower responses (Fox, 1984; Miller and Dexter, 1988; Pitt, 1995; Pitt and Samuel, 1993). In the categorization of syllable-final sounds, however, lexical influences get weaker in slower responses (McQueen, 1991; Pitt and Samuel, 1993). TRACE predicts that lexical involvement in both positions should build up over time because, in this model, feedback increases over time. TRACE is therefore challenged by the syllable-final results.

In the present study, categorization of both initial and final fricatives was tested. In most previous studies only one target position was tested; where both positions have been examined, position has tended to be a between-subjects factor (e.g., Pitt and Samuel, 1993). We therefore asked the same subjects to categorize both syllable-initial and syllable-final sounds, and examined the time course of their categorization performance. TRACE predicts that there should be lexical effects both initially and finally. Listeners should tend to label sounds in a W–NW series consistent with the word endpoint, but consistent with the other endpoint in the matched nonword–word (NW–W) series. Crucially, TRACE also predicts that lexical effects should build up over time. Simulations of this prediction in syllable-initial and syllable-final position are given, respectively, in (McClelland and Elman, 1986; McClelland, 1987). To examine this time-course prediction, we split listeners' responses into fast, medium and slow reaction time (RT) ranges (Fox, 1984; McQueen, 1991; Miller and Dexter, 1988; Newman et al., 1997).

## 2. Experiment 1

### 2.1. Method

#### 2.1.1. Subjects

Twenty-four student volunteers were paid to take part. All listeners were native speakers of Dutch, with no known hearing disorders.

#### 2.1.2. Materials

There were two sets of monosyllabic items. In one set, the initial consonant was a fricative varying in place of articulation from [f] to [s]. The other set used an acoustically different syllable-final [f]–[s] series. Each set consisted of a W–NW series, in which the [f] endpoint was a Dutch word and the [s] endpoint was not a Dutch word, and a NW–W series, where only the [s] endpoint was a word. Thus, for the initial-position set, the series were *flauw* (dull) – *slauw* and *flaap* – *slaap* (sleep), and for the final-position set they were *maf* (silly) – *mas* and *jaf* – *jas* (coat).

The initial [f]–[s] endpoints were based on natural tokens of *flaap* and *slaap*, recorded by a native speaker of Dutch in a sound-attenuated booth, and then digitized at 16 kHz. The two frication noises were extracted and adjusted to be of equal length (180 ms, by removal of some steady-state noise in the longer original). A 15-step series was made by adding the amplitudes of the two waveforms sample by sample in different proportions (McQueen, 1991). The proportions were equally spaced in 15 steps from 0 to 1.0 and were added pair-wise so as to sum to 1.0. The steps were then spliced onto the contexts *laap* (458 ms) and *lauw* (448 ms), taken from utterances of the nonwords *thlaap* and *thlauw*. The other set was made in the same way, based on [s] from *mas* and [f] from *maf*. The 15 steps of this series (all 252 ms) were spliced

onto the contexts *ma* (264 ms) and *ja* (277 ms), taken from utterances of *math* and *jath*.

A preliminary version of this experiment was run using these materials (otherwise it was the same as the experiment described here). No lexical biases were found for either the syllable-initial or the syllable-final fricatives. In both target positions, lexical effects tend to be stronger when the speech materials are degraded (Burton et al., 1989; McQueen, 1991; Pitt and Samuel, 1993). Listeners tend to rely on the speech signal in phonetic categorization, rather than stored knowledge, unless the signal is impoverished. The stimuli were therefore degraded by low-pass filtering at 3 kHz, following the procedure that has been used to degrade an English [s]–[ʃ] series (McQueen, 1991). The filter cut-off of 3 kHz was chosen so as to degrade the stimuli severely, in order to increase listeners' reliance on lexical knowledge. Although most of the frication noise for both [f] and [s] was removed, spectral analysis showed that some low-frequency noise remained for all members of both the syllable-initial and the syllable-final series. Eight steps from each series were then selected on the basis of pilot listening tests for use in the main experiment: the two endpoints plus original steps 2, 5, 7, 8, 10 and 14 from both series. Hereafter, these will be referred to as steps 1–8 in both positions.

#### 2.1.3. Procedure

Each of the eight steps in each of the four series was presented 16 times, in two lists, one consisting of the 256 initial-fricative items, the other of the 256 final-fricative items. Each list was pseudo-randomly ordered so that all items were spread evenly throughout the lists. Half of the subjects heard the initial items first; half heard the final items first. Subjects were tested in groups of up to three in separate carrels in a quiet room. The stimuli were presented once every 2 s at a comfortable level over headphones. Subjects were asked to decide whether the initial (or final) sound of each token was [f] or [s] and to press one of two labeled buttons "F" or "S". They were asked to respond on every trial, as fast and as accurately as possible. RTs were measured from item onset.

### 2.2. Results and discussion

Five subjects failed to distinguish between the endpoints of at least one of the series (the categorization functions were flat). This suggests that there was not sufficient information remaining in the low-pass filtered frication noises for all listeners to be able to distinguish between them. Only the data from the remaining 19 subjects were analyzed. Analyses of variance (ANOVAs) were performed on the overall proportion of [f] responses per subject in each series, collapsing over all eight steps. The mean proportions of [f] responses are plotted for each series and for each position in Fig. 1.

There were strong lexical effects spread over the series: listeners tended to identify the fricatives in a lexically-consistent manner (more [f] responses to the *flauw–slauw* and *maf–mas* series than to the *flaap–slaap* and *jaf–jas* series, respectively). These lexical effects were reliable ($F(1,18) = 26.9$, $p < 0.001$). In neither analysis was there an effect of position (initial versus final) nor an interaction of position with the lexical effect. The tendency to identify the fricatives in a way that was consistent with the word endpoints was equally strong for initial and final fricatives.

#### 2.2.1. Reaction time analyses

The mean and standard deviation for each subject's responses to each stimulus step (separately for initial and final fricatives) were computed, and then each individual RT within each of these subsets was translated into a $z$ score (Newman et al., 1997). Scores of 0.43 and −0.43 divide the $z$-score distribution into three equal portions. Responses with $z$ scores $< -0.43$ were fast and those with $z$ scores $> 0.43$ were slow; the remainder were medium. Mean RTs in each RT range were: fast, 741 ms; medium, 922 ms; and slow, 1210 ms. The mean proportions of [f] responses are plotted for each series in each RT range for each position in Fig. 2.

In ANOVAs based on the overall proportion of [f] responses per subject in each series in each RT range, there was a significant interaction of the lexical effect with RT range ($F(2,36) = 41.2$, $p < 0.001$): The lexical effect was significant in the

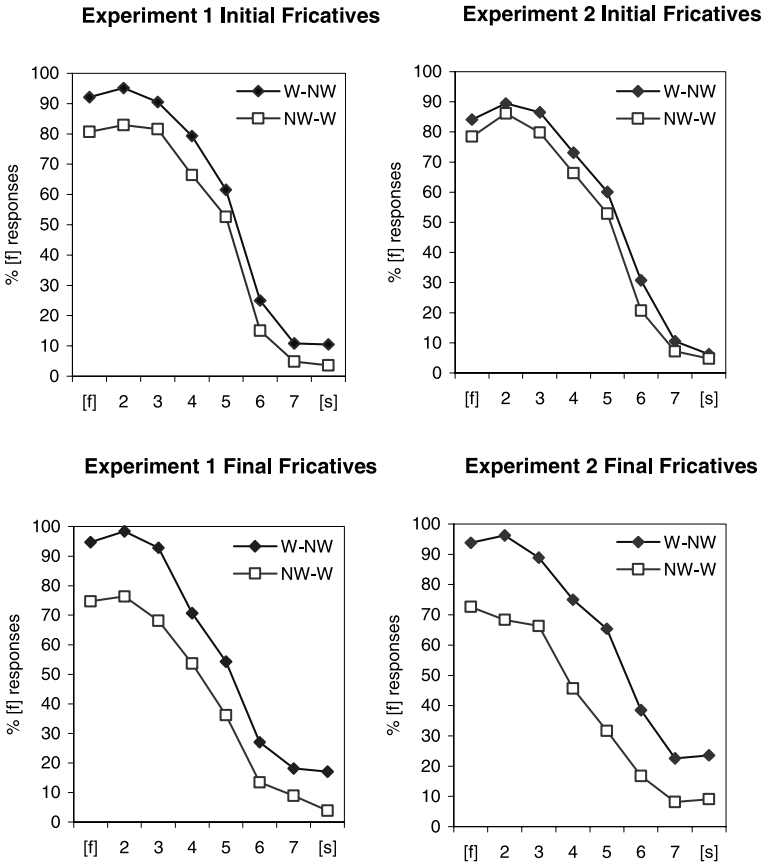


Fig. 1. Mean proportion of [f] responses to the W–NW and NW–W series, for syllable-initial (upper panels) and syllable-final categorization (lower panels). The data for Experiment 1 are given in the left panels; those for Experiment 2 in the right panels.

fast range ($F(1,18) = 57.8$, $p < 0.001$) and the medium range ($F(1,18) = 4.9$, $p < 0.05$), but not in the slow range ($F < 1$). There were no statistically significant differences between initial and final position in any of these analyses. In both cases, lexical effects became smaller in the slower responses (initial fricatives: fast, $F(1,18) = 24.6$, $p < 0.001$; medium, $F(1,18) = 2.3$, $p > 0.05$; slow, $F < 1$; final fricatives: fast, $F(1,18) = 31.8$, $p < 0.001$; medium, $F(1,18) = 3.5$, $p > 0.05$; slow, $F < 1$).

These results contradict the TRACE prediction that lexical effects should build up over time. While the results for syllable-final categorization are consistent with earlier findings, those for syllable-initial categorization are inconsistent with previous studies which have shown a build-up of lexical effects in slower responses. It is possible, however, that the listeners in Experiment 1 might have responded relatively slowly to the degraded fricatives. This could explain why lexical effects were found even for the fastest responses to the initial-position sounds. In Experiment 2, therefore, listeners were put under severe time-pressure (they had to try to respond within 500 ms of fricative offset). We predicted that these conditions would be more likely to yield very fast nonlexical responses, particularly to the syllable-initial items. We reasoned that if our failure in Experiment 1 to observe the build-up of lexical effects over time predicted by TRACE was because that build-up had occurred prior to responses being made even in the fastest RT range, it ought to be possible to

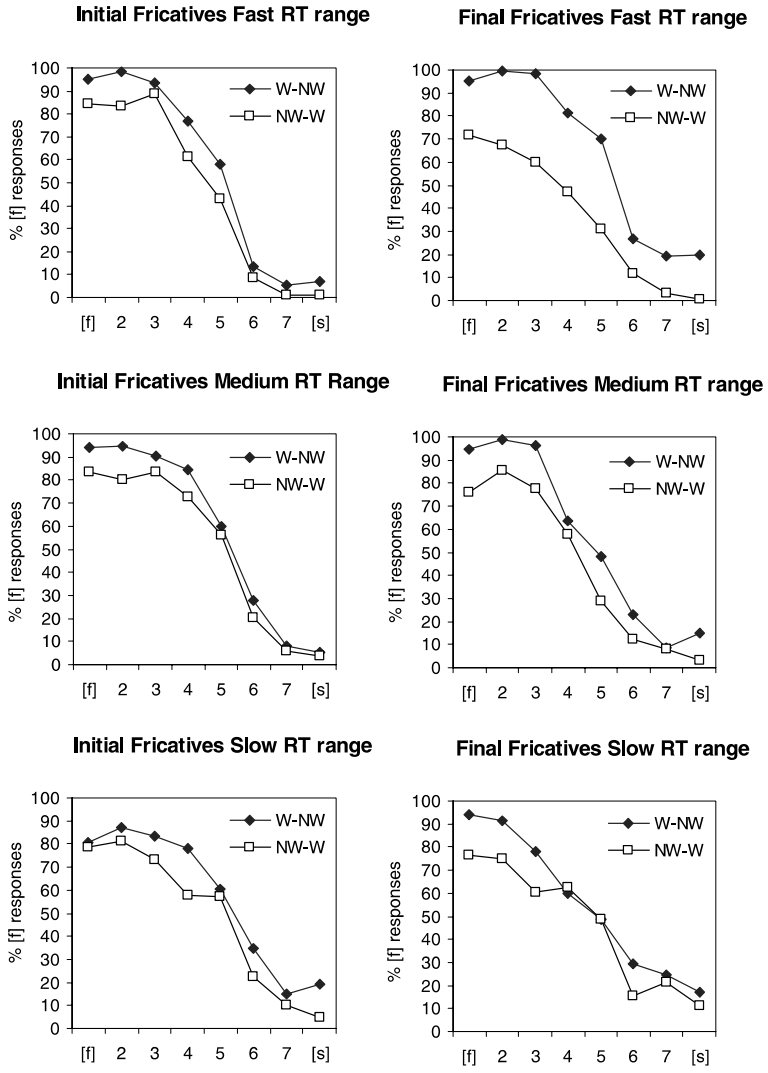


Fig. 2. Experiment 1: mean proportion of [f] responses to the W–NW and NW–W series, in the fast (upper panels), medium (middle panels) and slow RT ranges (lower panels), for syllable-initial (left panels) and syllable-final categorization (right panels).

observe that build-up if listeners are put under pressure to respond very fast.

## 3. Experiment 2

### 3.1. Method

#### 3.1.1. Subjects

Twenty-four student volunteers were paid to take part. All listeners were native speakers of Dutch, with no known hearing disorders. None had taken part in Experiment 1.

#### 3.1.2. Materials and procedure

The materials were the same as those used in Experiment 1. The procedure was also identical, except that a tone (1 kHz, 100 ms) was presented 500 ms after offset of the to-be-categorized fricative (whether it was initial or final). Subjects were asked to try on every trial to respond before they heard the tone.

### 3.2. Results and discussion

The data from 11 subjects were excluded because the extreme time pressure led them either to respond with considerable variability or to fail systematically to distinguish even between the endpoint tokens; only those subjects who were able to identify the endpoints with reasonable accuracy were included in the analysis. The time-pressure manipulation worked for these 13 subjects: mean RT was reduced from 958 (in Experiment 1) to 737 ms. The mean proportions of [f] responses are plotted for each series and for each position in Fig. 1.

There were again strong and reliable lexical effects across the series ($F(1,12) = 29.9$, $p < 0.001$). Although there was no main effect of position, there was an interaction ($F(1,12) = 11.6$, $p < 0.001$): lexical effects were stronger for the final fricatives. Separate ANOVAs by position in both types of analysis confirmed however that the lexical effect was reliable in both positions (initial: $F(1,12) = 17.8$, $p < 0.005$; final: $F(1,12) = 21.2$, $p < 0.001$).

#### 3.2.1. Reaction time analyses

The data were split into RT ranges in the same way as in Experiment 1. Mean RTs after this split were: fast, 594 ms; medium, 721 ms; and slow, 895 ms. The mean proportions of [f] responses are plotted for each series in each RT range for each position in Fig. 3.

There was again a significant interaction of the lexical effect with RT range (overall proportion, $F(2,24) = 17.7$, $p < 0.001$). The lexical effect was significant in the fast range ($F(1,12) = 52.6$, $p < 0.001$), and, although the effect was statistically larger in the final fricatives than in the initial fricatives ($F(1,12) = 16.9$, $p < 0.005$), it was significant in both subanalyses by position (initial: $F(1,12) = 13.3$, $p < 0.005$; final: $F(1,12) = 38.5$, $p < 0.001$). The lexical effect was also significant in the medium range ($F(1,12) = 8.8$, $p < 0.05$), but, although the interaction with position was not significant, the effect was significant only in the final-position subanalysis (initial: $F(1,12) = 3.8$, $p > 0.05$; final: $F(1,12) = 7.4$, $p < 0.05$). There were no reliable lexical effects in the slow range (overall: $F < 1$; initial: $F < 1$; final: $F(1,12) = 1.6$, $p > 0.05$). As in Experiment 1, lexical effects diminished in listeners' slower responses. Even under strong time pressure, the fastest responses to both initial and final fricatives were influenced by lexical information, but, contrary to TRACE's prediction, this effect died away rather than built up over time.

## 4. General discussion

In two categorization experiments with Dutch materials and Dutch listeners, in one of which listeners were placed under severe time pressure, significant lexical effects were observed for both initial and final fricatives: Throughout the series there were more [f] responses in the W–NW than in the NW–W series. Lexical involvement was strongest in the listeners' fastest responses but tended to disappear in their slowest responses.

The results with final fricatives mirror those found in previous studies (McQueen, 1991; Pitt and Samuel, 1993). It appears that even with degraded word-final fricatives, lexical knowledge is used in phonemic decision-making only within a limited time frame. These results challenge TRACE, which predicts that lexical effects should build up over time; as word activation rises, the amount of top-down facilitation from word to phoneme nodes should increase, and so too should the lexical bias in the categorization responses. One might argue that for word-final sounds, lexical activation may have already reached asymptote even in the fastest RT range, and thus perhaps that feedback may also have peaked soon after word offset. But the TRACE simulations reported in (McClelland, 1987) show that the activation levels of phoneme nodes continue to diverge well after word offset. It takes time in the model for lexical feedback to influence the activation of the phonemes nodes, gradually increasing the activation of the lexically-consistent interpretation of the ambiguous sound, and gradually decreasing the activation of the lexically-inconsistent alternative.

The initial fricative results contradict earlier findings (Fox, 1984; Miller and Dexter, 1988; Pitt, 1995; Pitt and Samuel, 1993). In the earlier studies, the strongest lexical effects appeared in this target

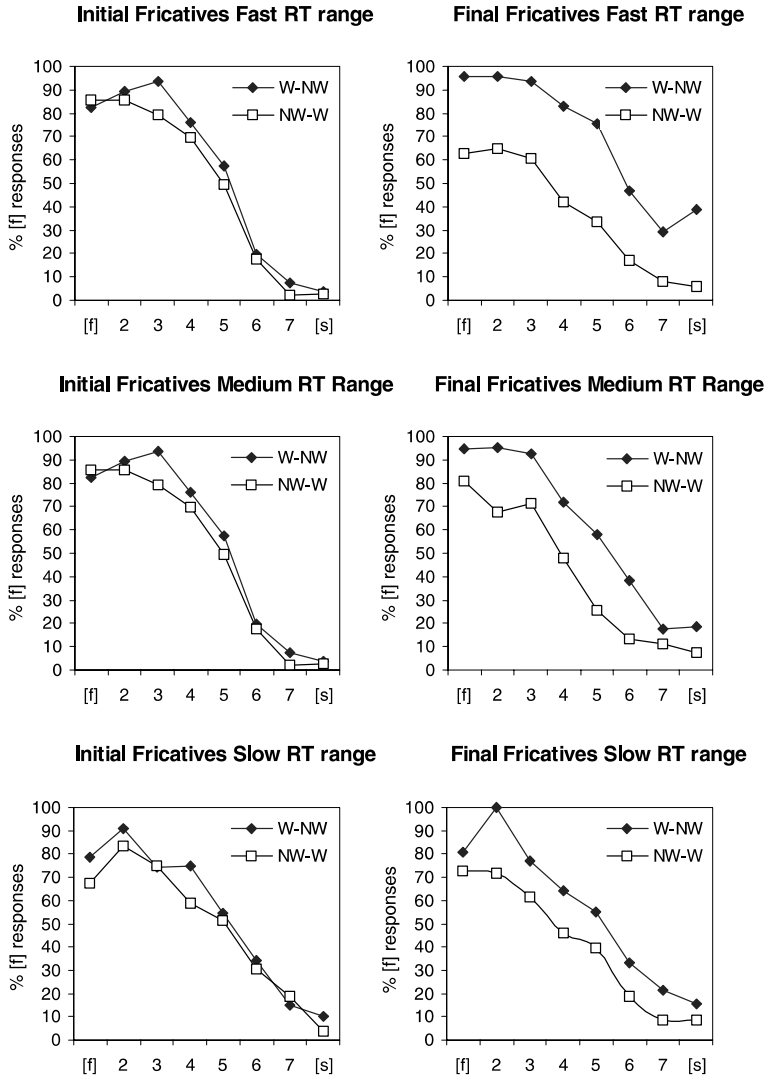


Fig. 3. Experiment 2: mean proportion of [f] responses to the W–NW and NW–W series, in the fast (upper panels), medium (middle panels) and slow RT ranges (lower panels), for syllable-initial (left panels) and syllable-final categorization (right panels).

position in the slowest responses. The explanation for this pattern was that very fast responses to initial segments can be initiated before the word is strongly activated and hence before word activation can influence decisions. In our study, the low-pass filtered initial fricative series appears to have been sufficiently degraded to prevent such rapid nonlexical responses being made. Even under severe time pressure, listeners appear to have waited until they had enough information to identify the context; lexical biases were thus seen in the fastest responses. Interestingly, although listeners could not ignore lexical information in their fastest responses to the initial fricatives, they tended not to use it in their slower responses. These results again challenge the TRACE prediction that lexical effects should build up over time.

The results are, however, consistent with the view, incorporated in the Merge model, that there is no feedback from the lexicon to prelexical levels.

One might argue that since Merge is an activation-based model, like TRACE, it too should predict that lexical effects should increase over time. This is not true. Let us assume that even in Experiment 2, the fastest responses reflect a stage of processing after any initial build-up of lexical activation. If we grant this to the Merge model, then we should grant it to TRACE. As we have already seen, however, TRACE predicts that the effects of feedback will continue to build up until well after stimulus offset. The reason for this is the feedback loop itself: as a word becomes more activated, there will be stronger top-down feedback to its constituent phonemes, they in turn will activate the word more, and the feedback will continue to strengthen. Positive feedback thus generates a stronger lexical bias on slower phoneme decisions than on faster decisions.

In contrast, Merge has no such feedback loop. Since the model only has feedforward connections, a decision node can not enter into a positive reinforcement loop with the prelexical and lexical representations with which it is connected. Nor can a word node enter into such a loop with its constituent phonemes at the prelexical level. This means that once a word representation has reached as strong a level of activation as it can given the available input (through a combination of bottom-up support, bottom-up inhibition and competition with other activated words) it can then increase in activation no more through the operation of positive feedback. This in turn means that the lexical bias (the influence of the word's activation on the decision nodes) will also reach asymptote. Merge thus does not predict that lexical effects should gradually increase in magnitude.

Our data are therefore problematic for models like TRACE which have perceptual feedback, that is, feedback from any given activated lexical representation which modulates the prelexical analysis of the component sounds of that word. Another problem with perceptual feedback is that it acts to distort information in the speech signal (Massaro, 1989; Norris et al., 2000). As the lexicon boosts the activation of a lexically-consistent phoneme node in TRACE, and the activation of the lexically-inconsistent node is being penalized by competition between the phoneme nodes, the representation of what was actually present in the input is being overwritten. It is thus hard to see how the ambiguity can, as it were, re-assert itself in slower responses, even if feedback is then switched off. In contrast, in Merge, if flow of information from the lexicon to the decision nodes is switched off later in processing, the ambiguity that was present in the signal can re-assert itself at the decision stage because it remains represented at the prelexical level (since the activation of the prelexical nodes does not die away immediately after stimulus offset, they act as a kind of memory buffer for the speech which has just been heard). A slow decision in Merge can thus be based on an accurate description of the signal, even though a lexically-biased response could have been made to the same stimulus if that response had been initiated earlier.

The tendency of perceptual feedback to distort the representation of the perceptual world is one strong argument against this kind of feedback, and thus against TRACE. A second theoretical argument against perceptual feedback is that it cannot benefit word recognition, and therefore has no function to serve in normal listening (Norris et al., 2000). Feedback can never improve recognition of a given word at the time that word is heard (if the prelexical level operates optimally, the same word will be recognized whether there is feedback or not). The perceptual feedback in TRACE is also challenged by data on compensation for coarticulation between fricatives and stops (Pitt and McQueen, 1998). If feedback is biasing the activation of fricative nodes in TRACE, it should in turn bias the interpretation of following stops through the operation of the mechanism in TRACE which compensates for fricative-stop coarticulation (Elman and McClelland, 1986, 1988). Pitt and McQueen, however, observed a dissociation: there was lexical involvement in fricative identification but not in stop identification.

The present data, those of Pitt and McQueen (1998), and the two theoretical arguments against feedback mentioned above challenge the view that there is perceptual feedback in the speech recognition system. They do not, however, challenge the other two types of feedback which we defined in the introduction: attentional feedback (where higher-level processing directs attention to particular

features of the input) and feedback for learning (where higher-level knowledge is used to retune perceptual categories). Our recent research (McQueen et al., 2001) suggests that listeners do indeed use lexical knowledge in retuning their phonetic categories. Dutch listeners tended to label an ambiguous fricative (midway between normal [f] and normal [s]) as [f] if they had been exposed to that sound in lexical contexts which were predictive of [f] (e.g., [kara?], based on the Dutch word *karaf*, carafe, where [karas] is not a Dutch word). Other listeners, who were exposed to the ambiguous fricative in [s]-biased lexical contexts (e.g., [karka?], based on *karkas*, carcase, where [karkaf] is a nonword), tended to label the same ambiguous sound as [s].

While further research is required to tie down the exact nature of this perceptual learning effect (e.g., whether it has an attentional component), it appears that listeners can use lexical knowledge to adjust their prelexical representations when they encounter speech which mismatches with normal perceptual categories. Interestingly, recent evidence which has been proposed to offer support for perceptual feedback (Samuel, 1997, 2001) has come from studies which had a learning component (selective adaptation experiments). The current weight of evidence thus suggests that while there is feedback for learning in spoken word recognition, there is no perceptual feedback.

A critical difference between feedback for learning and perceptual feedback is that only the former can benefit word recognition. While perceptual feedback can not improve recognition of a given word at the time that word is heard, longer-term learning can help the recognition system adjust to unusual speech (e.g., the speech of someone with an unfamiliar dialect). Adjustments to an unusual fricative, for example, would generalize to other words, in the sense that, after exposure to one set of words, recognition of other words containing the same fricative sound would improve. Perceptual feedback cannot benefit word recognition without a learning mechanism which allows generalization to the processing of other words. Another difference between these two types of feedback concerns when they act. The perceptual feedback in TRACE operates all the time, as every word is heard, and irrespective of the goodness of fit between the current speech signal and perceptual categories. In contrast, it seems likely that lexical retuning of phonetic categories will only occur when there is a consistent mismatch between the signal and stored knowledge.

One argument that has been made in favor of perceptual feedback is that there are widespread efferent neural connections in the brain (see the commentaries on Norris et al., 2000). The existence of these backprojections, however, does not mean that there is always feedback from each word to each of its constituent phonemes. It is essential to establish what function these connections serve, and that demands careful definition of the term "feedback". Our current, speculative position is that the backprojections play a role in perceptual learning, and that they may also play a role in attentional control, but that they do not operate such that the lexicon exerts an obligatory and immediate influence on the interpretation of every segment in speech.

In this article, we have discussed the information-processing architecture of the spoken word recognition system. We have argued that information in the speech signal is used at a prelexical level to generate abstract descriptions of spoken utterances. These descriptions are not purely segmental: In languages where lexical stress information can profitably be used to constrain lexical access, suprasegmental information is also extracted at the prelexical level. Information then flows up to the lexical level, where candidate words which are consistent with the input are activated. The computation of goodness of fit of these candidate words appears to use both bottom-up facilitation and bottom-up inhibition. Any incoming quantum of information increases activation of matching words, and decreases activation of words it mismatches. Information then flows laterally, as candidate words compete directly with each other via inhibitory connections. The presence of a more highly activated competitor further inhibits a word already inhibited by bottom-up mismatch. We have also argued that top-down flow of information from the lexical level back to the prelexical level during word recognition may only occur when it can facilitate perceptual learning, that is,

when it could improve word recognition on subsequent encounters with the same kind of speech.

## Acknowledgements

We would like to thank Pieter Meima for running these experiments, and Keith Kluender for helpful comments on a previous version of this paper.

## References

Burton, M.W., Baum, S.R., Blumstein, S.E., 1989. Lexical effects on the phonetic categorization of speech: The role of acoustic structure. Journal of Experimental Psychology: Human Perception and Performance 15, 567–575.

Connine, C.M., Clifton, C., 1987. Interactive use of lexical information in speech perception. Journal of Experimental Psychology: Human Perception and Performance 13, 291–299.

Connine, C.M., Titone, D., Deelman, T., Blasko, D., 1997. Similarity mapping in spoken word recognition. Journal of Memory and Language 37, 463–480.

Cutler, A., 1986. *Forbear* is a homophone: Lexical prosody does not constrain lexical access. Language and Speech 29, 201–220.

Cutler, A., van Donselaar, W., 2001. *Voornaam* is not (really) a homophone: Lexical prosody and lexical access in Dutch. Language and Speech 44, 171–195.

Cutler, A., Mehler, J., Norris, D., Seguí, J., 1987. Phoneme identification and the lexicon. Cognitive Psychology 19, 141–177.

Cutler, A., Dahan, D., van Donselaar, W., 1997. Prosody in the comprehension of spoken language: A literature review. Language and Speech 40, 141–201.

Elman, J.L., McClelland, J.L., 1986. Exploiting lawful variability in the speech wave. In: Perkell, J.S., Klatt, D.H. (Eds.), Invariance and Variability of Speech Processes. Erlbaum, Hillsdale, NJ, pp. 360–380.

Elman, J.L., McClelland, J.L., 1988. Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. Journal of Memory and Language 27, 143–165.

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

Frauenfelder, U.H., Floccia, C., 1998. The recognition of spoken words. In: Friederici, A. (Ed.), Language Comprehension: A Biological Perspective. Springer, Berlin, pp. 1–40.

Frauenfelder, U.H., Scholten, M., Content, A., 2001. Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. Language and Cognitive Processes 16, 583–607.

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

Gaskell, M.G., Marslen-Wilson, W.D., 1997. Integrating form and meaning: A distributed model of speech perception. Language and Cognitive Processes 12, 613–656.

Luce, P.A., Goldinger, S.D., Auer, E.T., Vitevitch, M.S., 2000. Phonetic priming, neighborhood activation, and PARSYN. Perception & Psychophysics 62, 615–625.

Marslen-Wilson, W., Warren, P., 1994. Levels of perceptual representation and process in lexical access: Words, phonemes, and features. Psychological Review 101, 653–675.

Massaro, D.W., 1989. Testing between the TRACE model and the fuzzy logical model of speech perception. Cognitive Psychology 21, 398–421.

McClelland, J.L., 1987. The case for interactionism in language processing. In: Coltheart, M. (Ed.), Attention and Performance XII: The Psychology of Reading. Erlbaum, Hillsdale, NJ, pp. 3–36.

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

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

McQueen, J.M., Norris, D., Cutler, A., 1999. Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. Journal of Experimental Psychology: Human Perception and Performance 25, 1363–1389.

McQueen, J.M., Norris, D., Cutler, A., 2001. Can lexical knowledge modulate prelexical representations over time? In: Proceedings of the Workshop on Speech Recognition as Pattern Classification. MPI for Psycholinguistics, Nijmegen, pp. 9–14.

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

Newman, R.S., Sawusch, J.R., Luce, P.A., 1997. Lexical neighborhood effects in phonetic processing. Journal of Experimental Psychology: Human Perception and Performance 23, 873–889.

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

Norris, D., McQueen, J.M., Cutler, A., 2000. Merging information in speech recognition: Feedback is never necessary. Behavioral and Brain Sciences 23, 299–370.

Pitt, M.A., 1995. The locus of the lexical shift in phoneme identification. Journal of Experimental Psychology: Learning, Memory, and Cognition 21, 1037–1052.

Pitt, M.A., McQueen, J.M., 1998. Is compensation for coarticulation mediated by the lexicon? Journal of Memory and Language 39, 347–370.

Pitt, M.A., Samuel, A.G., 1993. An empirical and meta-analytic evaluation of the phoneme identification task. Journal of

Experimental Psychology: Human Perception and Performance 19, 699–725.

Pylyshyn, Z., 1999. Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. Behavioral and Brain Sciences 22, 341–423.

Rubin, P., Turvey, M.T., van Gelder, P., 1976. Initial phonemes are detected faster in spoken words than in nonwords. Perception & Psychophysics 19, 394–398.

Samuel, A.G., 1981. Phonemic restoration: Insights from a new methodology. Journal of Experimental Psychology: General 110, 474–494.

Samuel, A.G., 1996. Does lexical information influence the perceptual restoration of phonemes? Journal of Experimental Psychology: General 125, 28–51.

Samuel, A.G., 1997. Lexical activation produces potent phonemic percepts. Cognitive Psychology 32, 97–127.

Samuel, A.G., 2001. Knowing a word affects the fundamental perception of the sounds within it. Psychological Science 12, 348–351.

Soto-Faraco, S., Sebastián-Gallés, N., Cutler, A., 2001. Segmental and suprasegmental mismatch in lexical access. Journal of Memory and Language 45, 412–432.