Lexical Influences on the Perceptual Categorization of French Stops

W. SERNICLAES
R. BEECKMANS †
M. RADEAU

Abstract

Lexical effects on speech perception are not very reliable and they have been shown to depend on various factors among which word length. In the current models of phonemic decision, lexical effects are conceived as arising from top-down processing, with or without feedback, depending on the model. Lexical effects tend to be stronger in longer words, which can be ascribed to an increase in the amount of lexical evidence. The present study was aimed at collecting further evidence on this point. The existence of lexical effects was confirmed in a series of two experiments on voicing identification in French initial stops. The effects were present for stops in monosyllables and polysyllables whereas they were almost absent in bisyllables. We tentatively explain the U-shaped relationship between lexical evidence and phonemic identification by two different mechanisms which would be both weakly effective with moderate amounts of lexical evidence (in bisyllables). With fairly large amounts of lexical evidence (in polysyllables) the lexical effect would be due to the fairly complex top-down processes postulated in the literature. With low amounts of lexical evidence (in monosyllables), a much simpler mechanism based on a re-analysis of the acoustic input would be at work.

Keywords: Phoneme perception; Lexical effects; Voicing; French stops

1. Introduction

Phonetic features are elementary distinctions in the production and perception of speech. Feature categorization is achieved by fairly complex processes which allow the listener to integrate multiple acoustic cues into a

single phonetic percept (Repp, 1982). Further, perception of phonetic features not only depends on acoustic factors but also on contextual features (Serniclaes & Wajskop, 1992). Moreover, lexical factors seem to affect feature perception (Ganong, 1980). When asked to identify phonetic features, listeners tend to favor lexical answers in case of acoustic ambiguity. For instance, when asked to identify initial stops subjects provided more voiceless responses with a nonword-word *dask-task* contrast than with a word-nonword *dash-tash* contrast. The phoneme boundary was thus shifted towards the word-end of the continuum. This effect is generally referred to as a "Lexical Identification Shift". Lexical influences have also been demonstrated for phoneme restoration (Samuel, 1996), phoneme monitoring (Rubin, Turvey & van Gelder, 1976; Frauenfelder, Segui & Dijkstra, 1990) and detection of pronunciation errors (Cole, 1973; Cole, Jakimik, & Cooper, 1978).

Phoneme position in word affects the amount of lexical evidence, taking for granted that the analysis of the word stimulus proceeds from the onset to the end, as in the original version of the Cohort model (Marslen-Wilson & Welsh, 1978), and as demonstrated at least for the slow speaking rates usually used in experiments devoted to words presented in isolation (Radeau, Morais, Mousty & Bertelson, 2000). Dependency of lexical effects on phoneme position in word was evidenced for different phoneme decision tasks. In phoneme monitoring, lexical influences are more robust for final vs. initial syllables (Frauenfelder et al., 1990; Pitt & Samuel, 1990). Lexical influences have also been shown to increase for late syllables with two other tasks. The so-called "phoneme restoration effect", i.e. the fact that phonemes are still perceived when the corresponding segment in the acoustic signal is replaced by noise (Warren, 1970), is stronger for words versus nonwords and the difference increases for late syllables in the word (Samuel, 1996). Detection of pronunciation errors is more difficult for non-initial phonemes or syllables (Cole, 1973; Cole et al., 1978). Finally, a more recent study showed that the lexical shift in feature identification is higher for polysyllables than for monosyllables (Pitt & Samuel, 2006).

In summary, evidence collected with different paradigms, i.e. phoneme monitoring, phoneme restoration, detection of pronunciation errors and phoneme identification on word-nonword continua, suggest that lexical influences on phoneme perception increase as a function of the amount of available lexical evidence.

Models: interactive vs. autonomous

Lexical effects were first taken as strong support for interactive theories of speech recognition. An example of interactive account is that proposed by TRACE (McClelland & Elman, 1986) which assumes top-down facilitatory connections from the word to the phoneme level and from the phoneme level to the feature level. Activation of word processing units (word nodes) by their component phonemes would, after inhibitory competition between words,

give rise to feedback activation of phoneme nodes. The latter would in turn influence feature recognition by similar inhibition and feedback mechanisms. Lexical effects would thus result from lexical facilitation of acoustic-phonetic decoding.

However, another class of models explains lexical effects without lexical feedback. In an autonomous model such as Race (Cutler & Norris, 1979), phoneme identification is performed in parallel along two different routes, phonemic or lexical. Phoneme identification along the phonemic route is based on phonetic information, whereas the lexical route is only available after word recognition since it relies upon phonological descriptions stored in the lexicon. According to this model, both routes converge to a decision node that only takes account of the most rapid one. Although the lexical route is less direct, it is more likely to win the race in some conditions. The lexical route will tend to be faster for target phonemes located closer to the end of the word, especially when phonetic information is ambiguous, which increases processing time. Lexical effects can then arise from the combination of two different outputs without any need for the processes delivering these outputs to influence each other.

Although both Race and TRACE can account for the early data on lexical perception, both fail to explain the wide variety of phenomena that have been evidenced over the years with the help of several different experimental tasks. Norris, McQueen and Cutler (2000) examined the limits of the two models and concluded that each of them is seriously challenged by some of the results published in the literature.

How are top-down effects regulated?

Some of these problems are related to the fact that in TRACE top-down effects bias lower-level decisions without taking account of the lower-level evidence. Unrestricted top-down bias in TRACE makes that this model cannot account for detection of mispronunciations and lack of inhibitory effects in nonwords.

Mispronunciations would not be detected with TRACE because, according to this model, unambiguous lower-level evidence can be overturned. For instance, a clearly mispronounced /d/ as in /dask/ would be perceived as /t/ under the influence of top-down lexical information. Unambiguous bottom-up information would be completely ignored if it contradicts the lexical evidence. This goes against everyday experience showing that mispronunciations are perceptible although experimental data show that they are not always detected (Cole, et al., 1978).

Lack of inhibitory effects in nonwords, together with their occurrence in words, refers to the absence of lexical effects on the detection of a deviant phoneme (e.g. on the detection of /t/ in the French nonword *vocabutaire*, which differs from the word *vocabulaire* by the target phoneme only, when matched with a nonword *socabutaire*; Frauenfelder et al., 1999). The latency

of phoneme detection should be longer for the nonword *vocabutaire* if the perception of /t/ was inhibited by the lexically induced perception of /l/ from the word *vocabulaire*. However, according to Norris et al. (2000, p.307) this would not raise a problem for a mechanism which does not inhibit the phonetic evidence (supporting /t/ here) if the latter were strong enough.

Detection of mispronunciations, as well as various phenomena such as lack of inhibition on nonwords, cannot be explained if lexical evidence were not constrained by the phonetic one. A model with top-down processes should thus specify suitable rules for integrating phonetic information. Norris et al. (2000) take this into account in the framework of an upgraded autonomous model. In a new model, called "Merge", phoneme recognition is affected by lexical top-down information, as in TRACE, but without phonemic feedback towards the lexical level, as in Race. Two different phoneme decision layers are considered, the lower level being immune to lexical top-down influences whereas the upper level is affected both by lexical effects and lower-level phonemic inputs. This architecture accounts for lexical effects on nonwords evidenced in different studies (Marslen-Wilson & Warren, 1994; Connine, Titone, Deelman & Blasko, 1997; McQueen, Norris & Cutler, 1999) and which raised a basic problem for Race because only segments in words can be processed along the lexical route with this model. Moreover it can avoid the TRACE's failure to cope with detection of mispronunciations and related phenomena through the incorporation of a "bottom-up" priority rule that prevents the phoneme decision nodes to be active in the absence of bottom-up support. However, further arguments in support of interactive models have been proposed (McClelland, J., Mirman, D., & Holt, 2006) but these arguments were in turn refuted (McQueen, Norris, & Cutler, 2006). Yet, recent neuroimaging evidence suggests that the pattern of neuroimaging activation is more in line with the interactive view: higher-level lexical information seems to directly influence the perception of incoming speech (Myers & Blumstein, 2008). In summary, it would seem that lexical information somehow affects speech perception, against Merge, but that the lexical influence can only operate within some limits, against TRACE.

The present study

In this paper, we raise the question of how precisely the lexical evidence is combined with the phonetic one. Trying to answer this question leads one to examine how lexical effects change as a function of the amount of available lexical evidence. For this purpose, we examined the combined effects of word length and phoneme position on lexical effects in phoneme identification by French listeners.

Phoneme position in word affects the amount of lexical evidence and joint evidence from several different tasks suggest that lexical effects on phoneme perception is larger for late syllables in the word (see above, detection of pronunciation errors: Cole et al., 1978; phoneme monitoring: Frauenfelder et

al., 1990; phoneme restoration effect: Samuel, 1996). However, the effect of phoneme position on phoneme categorization is not very reliable (see above, Pitt & Samuel, 1995). Word length is inversely correlated with the density of lexical neighborhood, the latter being defined as the number of words of the same length differing by one phoneme in any position (Luce, 1986; Newman, Sawusch, & Luce, 1997). Density is closely related to lexical evidence and has an effect on phoneme categorization although this depends on the phonetic context (Newman et al., 1997). While both phoneme position in word and word length have potential implications for phoneme categorization, the effects of each of these two are not very reliable. This can be understood by the fact that these factors affect the amount of lexical evidence only insofar they modify the amount of lexical competition irrespective of word length. Vitevitch, Stamer & Sereno (2008) showed that bisyllables with sparse neighborhoods were more accurately identified than those with dense neighborhoods. Therefore, combining word length and phoneme position modifications offers a potentially more reliable procedure for evidencing the effects of the of amount lexical evidence on phoneme categorization, i.e. the addition of two unreliable factors gives a better chance of observing the effect, at least if the factors are not completely redundant.

As the present study was conducted with French-speaking subjects and given the variability of lexical effects (Pitt & Samuel, 1993), the possibility of obtaining a lexical shift in French was first assessed (Preliminary Experiment). Most of the studies devoted to lexical effects on phoneme identification were indeed performed in English. The purpose of the Preliminary Experiment was to replicate these effects in French. In Experiment 1, lexical effects were compared for phonetically similar target phonemes in monosyllables vs. bisyllables. Lexical effects in polysyllables were investigated in Experiment 2.

2. Experiment 1. Monosyllables & Bisyllables

As already described in the Introduction, previous data suggest that lexical effects are somewhat more robust for late occurring phonemes. We therefore expected to find *stronger* lexical effects on phoneme identification in late positions. The aim of the present experiment was to test this hypothesis. We compared lexical effects in monosyllabic words, with a high density of lexical neighbors, vs. bisyllabic words whose lexical density is necessarily lower. In each case, the target phoneme was located at the beginning of the same syllable that was in either initial position (CVC words) or final position (CVCVC words). Such a joint modification of word length and phoneme position has the advantage of conflating two different lexical parameters, namely phoneme position in word and word length while keeping the stress pattern unchanged since stress always falls on the last syllable in French.

Moreover, choosing to locate the target phoneme in the last rather than in the first syllable is the best way to assess the effect of density of lexical neighborhood. Indeed, it is quite possible that for initial syllables in bisyllabic words the lexical neighbors taken into account by the perceptual system are those corresponding to the first embedded word-syllables rather than to the entire word-stimulus and are thus about the same in both stimuli (Vroomen & de Gelder, 1997).

2.1. Preliminary Experiment

As lexical effects have been evidenced for initial stops in English, but not in French, a preliminary experiment was run in order to rule out a possible language effect. The procedure was similar to the one previously used for English (e.g. Ganong, 1980). Two minimal voicing pairs between initial stops in CVC syllables were synthesized. In one pair, *dame-tame*, the lexical item (*dame* means "*lady*") began with a voiced stop (/d/), while in the other, *dasse-tasse* (*tasse* means "*cup*"), it was initiated with a voiceless stop (/t/). Participants were asked to identify the initial stop by choosing between D and T labels.

Two VOT continua (/dam-tam/ & /das-tas/), were created with a cascade/parallel formant synthesizer (Klatt, 1980; Bailly & Guerti, 1991). VOT ranged between – 40 and + 40 ms in 8 steps of 10 ms each and was centered on the French boundary which is located around 0 ms (Serniclaes, 1987). The stimuli were presented 10 times in an experimental series and identification responses were collected in a set of 15 French-speaking subjects (16 to 35 years of age).

Percent voiced responses for the /dam-tam/ was larger than for the /das--tas/ contrast (Figure 1).

As the dependent variable was a percentage, an arcsine transform was used for testing the effect with ANOVA (Scheffé, 1959). The difference between mean percentage of voiced responses for the two contrasts was significant in a repeated measures ANOVA with Pair (dam/tam vs. tas/das) as within--participants factor ($F(1,14) = 12.4$; $p<.01$). The phonetic boundary was shifted towards larger VOT values for dam/tam vs. das/tas. Boundary estimations were obtained by fitting Logistic functions and yielded values of 6.1 ms VOT for the das/tas contrast and 12.2 ms VOT for the dam/tam contrast. Logistic regression tests show that the difference between these boundaries was significant (Wald test; $p < .01$).

These results show that a lexical effect can be obtained for French stops in initial position. The magnitude of the lexical shift (6.1 ms) is certainly not smaller than those reported for voicing contrasts in English initial stops, for which the range of significant lexical shifts is between 2.2 ms and 5.2 ms (Serniclaes, Beeckmans & Radeau, 1995).
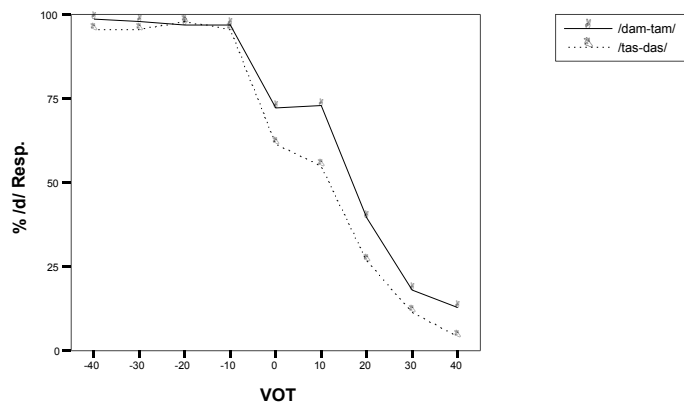
Figure 1: Results of the preliminary experiment. Percentage of voiced responses for stops in initial syllable position in two different word – nonword pairs. In /dam-tam/, /dam/ is a French word and /tam/ is a nonword. In /tas-das/, /tas/ is a French word and /das/ is a nonword.

2.2 Experiment 1. Method

Stimuli. The speech material used as source signals for building the stimulus continua was a set of French words and nonwords pronounced by a Belgian man native speaker of French. The complete list of source signals is presented in Table I. The items were digitized at a sampling rate of 48 kHz with the help of the Soundtools software and stored on computer disk. The stimuli were constructed with the help of the editing facilities provided by the Soundtools software. The words were either monosyllabic or bisyllabic. The mean syllabic rate was 1.74 per sec. for monosyllables and 2.64 per sec. for bisyllables. Six monosyllabic CVC words were used, three with an initial voiced consonant (*dame_, guerre_, gueule_* meaning *lady*, *war* and *face,* respectively) and three with an initial voiceless consonant (*tasse_, quel_, coeur_,* meaning *cup*, *who* and *heart,* respectively). The UP, indicated by the position underlined, is always located after the end of the word (see Table II for the phonetic transcription of the stimuli in IPA symbols).

For each of the six monosyllabic words there was a corresponding nonword with identical phonetic structure except for the voicing category of the initial stop. Each monosyllabic word and nonword also had a corresponding bisyllabic CVCVC word with the same final CVC part. These were: *madame, naguère, bégueule, potasse, nickel, moqueur* meaning *lady, formerly, prude, potash, nickel,* and *mocking* respectively. The UP is located just after the target phoneme except in one case (*naguère*) where it is located at the target. A voiced-voiceless continuum was constructed for each of the 12 word-nonword pairs.

Table I: Endpoints of the stimulus continua used in the two experiments are presented in the first column. Target phonemes are presented first followed by word and nonword continua endpoints. UP locations are indicated in bold. Lexical frequencies of the word stimuli of the two experiments (in log10 x frequency units) are presented in the second column. Densities of lexical neighborhood are presented in the third column. Stars indicate significant (p < .05) lexical effects.

Word Length	Word Frequency	Density of Lexical Neighborhood
Monosyllables in Exp. 1 **UP** is located after the end of the syllable *d/t: dame/tame & ta**ss**e/da**ss**e** *g/k: **gu**erre/**qu**erre & quel/guel** *g/k:**gu**eule/**qu**eule & **c**oeur/**gu**eur**	5.11 & 4.33 5.77 & 5.59 4.46 & 5.78	6/10 & 13/7 8/5 & 9/7 4/7 & 16/9
Bisyllables in Exp. 1 **UP** is located just after or at the target phoneme *mad**a**me /matame & pot**a**sse/podasse* *g/k: na**gu**ère /naquére& nick**e**l/niguel** *g/k: bégu**eu**le/béqueule & moqu**eu**r /mogueur*	5.04 & 3.15 4.46 & 3.26 2.62 & 3.85	0/0 & 5/1 0/0 & 1/0 0/0 & 0/0
Polysyllables in Exp. 2 **UP** is located just before the target phoneme *b/p: hurl**u**berlu/ hurluperlu & olnuberlu / olnuperlu** *d/t: hir**on**delle/ hirontelle & ounondelle / ounontelle* *g/k: élo**n**gation/ élonkation & ivongation / ivonkation* *b/p: acho**p**pement/ achobement & ouroppement / ourobement** *d/t: lév**i**tation / lévidation & fusitation / fusidation* *g/k: fun**i**culaire / funigulaire & namiculaire / namigulaire*	2.93 3.86 2.08 2.95 2.46 3.15	0/0 & 0/0 0/0 & 0/0 0/0 & 0/0 0/0 & 0/0 0/0 & 0/0 0/0 & 0/0

Lexical frequencies of the word stimuli, converted into log10 X frequency units, were assessed with Brulex (Content, Mousty & Radeau, 1990), are shown in Table I. In the two cases where two homophonic forms coexisted, the value taken into account was that of the most frequent form. As proposed by Newman et al. (1997), measurements of lexical density in number of neighbors differing only by the value of one of the phonemes were also performed for each endpoint. For instance for the word *quel*, the neighbors were *pelle, tel, bel, ..., calle, col ..., caisse, quête ...*, while for the nonword *guel*, they were *pelle, tel, bel, guêpe, ...galle, gueule, ...guerre, gaine....* For nonwords, the contrastive word in the pair (e.g. *quel*) was not counted as a neighbor. The lexical neighborhood densities are given in Table I for each word and nonword stimulus.

**Table II: VOT (Voice Onset time) and SI (Silent Interval during closure)
characteristics of the stimuli in Exps. 1 and 2. For the word stimuli, the location of the
Uniqueness Point (UP) is underlined. The first number in each cell corresponds to
voiced endpoint and the last one to the voiceless endpoint.**

continua endpoints (in IPA symbols)	VOT (ms)	SI (ms)
tas / das & dam / tam	-40 /+30	
kel / gel & ger / ker	-40/+40	
koer /goer & goel / koel	-40/+40	
potas / podas & madam / matam	0/+30	0/90
nikel / nigel & nager / naker	0/+30	0/80
mokoer / mogoer & begoel / bekoel	0/+40	0/120
irɔ̃del / irɔ̃tel & unɔ̃del / unɔ̃tel	0/+26	0/130
yrlyberly / yrlyperly & olnyberly / olnyperly	0/+10	0/82
elɔ̃gsjɔ̃ / elɔ̃kasjɔ̃ & ivɔ̃gasjɔ̃ / ivɔ̃kasjɔ̃	0/+14	0/83
levitasjɔ̃/ levidasjɔ̃ & fusitasjɔ̃ / fusidasjɔ̃	0/+40	0/56
a ʃopoemɑ̃ / a ʃoboemɑ̃ & uropoemɑ̃ / uroboemɑ̃	0/+18	0/99
fynikyler / fynigyler & namikyler / namigyler	0/+51	0/118

For the six CVCVC, the voicing continua were obtained by modifying the Voiceless Interval (VI), which corresponds to the sum of the closure silent interval (SI) and the VOT. The VI is a major cue for voicing perception in French medial stops (Saerens, Serniclaes & Beeckmans, 1989). The continuous periodic segment of the medial voiced stop was progressively replaced by the SI and VOT of its voiceless cognate. The SI and VOT values of the voiceless stops are given in Table II. Eleven SI and VOT values were chosen in such a way that the perceptual voicing boundary falls about in the middle of the continuum. Phonetic boundary position was assessed in a preliminary experiment. This gave three voiced- voiceless bisyllabic continua in which the voiceless endpoint corresponds to a word (e.g. /podas- potas/) or "voiceless word" continua.

The three other voiced-voiceless bisyllabic continua, for which the voiced endpoint corresponds to a word (e.g. /madam-matam/) or "voiced word" continua, were constructed with the acoustic cues already used for the "voiceless word" continua. The procedure was conceived in such a way as to minimize possible differences in the acoustic values of the voicing cues

between the two continua. For each continuum, the acoustic segment starting at the beginning of the medial stop closure and ending at the end of the following vocalic transition was then extracted from the stimuli corresponding to the "voiceless word" continuum (e.g. from the 11 stimuli of /podas-potas/ and each of the 11 segments was inserted in the corresponding "voiced word" (e.g. in /madam/). This gave three voiced- voiceless monosyllabic continua, one for each vocalic context, in which the voiced endpoint corresponds to a word (e.g. /madam-matam/) or "voiced word" continua.

A similar procedure was used for constructing the six CVC continua. However, instead of increasing the VI around the release (by concomitant modification of SI and VOT), the negative VOT of the original voiced stop was first progressively reduced. Positive VOT of the voiceless cognate was then progressively introduced after complete excision of the negative VOT. The VOT and SI values of the continua endpoints are given in Table II. A total of 11 VOT values were chosen in such a way that the perceptual voicing boundary falls about in the middle of the continuum. This gave three voiced-voiceless monosyllabic continua, one for each vocalic context, in which the voiceless endpoint corresponds to a word (e.g. /das-tas/) or "voiceless word" continuum. For each continuum, the acoustic segment starting at initial voice onset and ending at the end of the following vocalic transition was extracted from the stimuli corresponding to the "voiceless word" continuum (e.g. from the 11 stimuli of /das-tas/). Each of the 11 segments was then inserted in the corresponding "voiced word" (e.g. in /dam/). This gave three voiced-voiceless monosyllabic continua in which the voiced endpoint corresponds to a word (e.g. /dam-tam/) or "voiced word" continua. The VOT and SI values of the continua endpoints are given in Table II.

Participants. There were 20 French-speaking students, 11 women and 9 men aged between 18 and 25 years.

Procedure. The stimuli series were presented binaurally through headphones. Listeners were tested individually. They were asked to identify the first consonant as belonging to the /d,g/ set or to the /t,k/ set by using one of two keys on a computer-controlled response box. Stimuli were delivered at a fixed rate of one per 5 seconds. An initial block of 50 stimuli was used for practice and responses to these trials were not included in the data. The 132 stimuli (12 series X 11 stimuli) were presented in an experimental series in which each stimulus appeared 10 times. The series were presented in two sessions.

2.3 Experiment 1. Results and discussion

Figure 2 gives the percent voiced responses for word-nonword pairs in which the words contained a voiced stop vs. pairs in which the words contained a voiceless stop, either in monosyllabic stimuli (Figure 2a), or in the final syllable of bisyllabic stimuli (Figure 2b). As can be seen, voiced responding is larger for words with voiced target stops than for words with

voiceless target stops in the monosyllables whereas no consistent trend is apparent for the bisyllables.
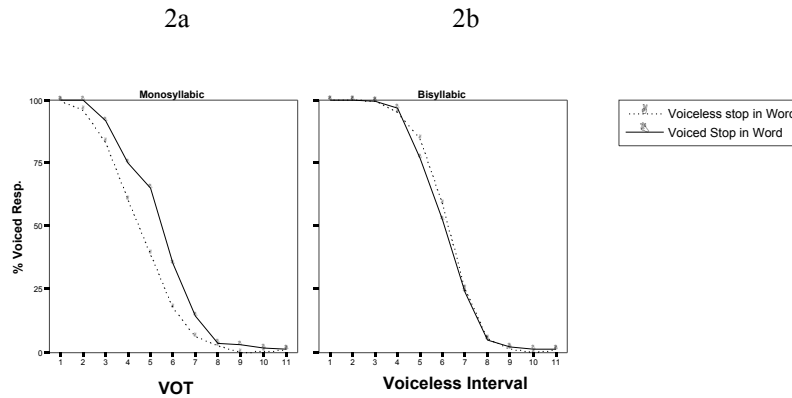
2a   2b

Figure 2: Results of Experiment 1. Mean percentages of voiced responses for the word-nonword monosyllabic continua (left) and for the bisyllabic ones (right). Responses for word-nonword continua for which an advantage for voiced responding was expected are indicated in continuous lines, those for which an advantage in voiceless responding was expected are indicated in broken lines.

Arcsine transforms of percentages were used for performing the tests. The whole data were first tested in a repeated measures ANOVA with Lexicality (voiced stop in word-voiceless in nonword context vs. voiced in nonword-voiceless in word context), Word Length (monosyllabic vs. bisyllabic) and Context (dame/madame & tassse/ potassse vs. guerre/naguère & quel/nickel vs. gueule/bégueule & coeur/moqueur) as within-participants factors.

Lexicality and the Lexicality X Word Length interaction were significant $(F(1,19) = 30.9, p<.001; 11.1, p<.01$, respectively), whereas the Lexicality X Context and Lexicality X Context X Word Length interactions were not significant $(F(2,38)= 2.84, p=.07; 1.31, p=.28)$. The lexical effect was thus reliable overall but depended on word length and was much smaller for the bisyllables (0.9 %, SE= 1.8) than for the monosyllables (7.1 %, SE= 1.6).

When tested separately for each Word Length, Lexicality was significant for the monosyllables but not for the bisyllables $(F(1,19)= 31.2, p<.001; F<1$, respectively). The Lexicality X Context was not significant for the monosyllables but it was marginally significant the bisyllables $(F<1; F(2,38)= 3.19, p=.05$; respectively). The Lexicality X Context interaction is due to idiosyncratic differences between the contexts and is not readily interpretable. Significance tests of Lexicality for each context are reported in Table I.

The absence of significant lexical effect in bisyllables was quite unexpected. Different studies suggested that lexical influences on speech

perception increase with word length (see Introduction: Cole et al., 1978, Frauenfelder et al., 1990) and, more importantly for our concern here, a more recent study showed that the lexical shift in feature identification is higher for polysyllables than for monosyllables (Pitt & Samuel, 2006). As previous studies suggest that the lexical shift increases with word length, we should have obtained an increase in lexical shift for bisyllables vs. monosyllables instead of the decrease which was effectively obtained. However, the increase in lexical shift was evidenced with differences between polysyllables and monosyllables in previous studies. The difference between our results with French speakers and those obtained in previous studies with English speakers might be due to language. The aim of the following experiment was to see whether a lexical shift is present in polysyllables for French speakers.

Finally, the *Dame/Tame* and *Dasse/Tasse* contrasts were used both in this experiment and in the Preliminary Experiment but they were generated with natural speech here and with synthetic speech in the Preliminary Experiment. This gave an opportunity to test a possible influence of stimulus type on lexical effects, as experiments in English suggest that lexical effects tend to diminish with better stimulus quality (Burton et al., 1989). However, the difference in lexical effects between the natural and synthetic versions of the D/T contrasts was not significant (t<1).

3. Experiment 2. Polysyllables

In this experiment, lexical effects on the recognition of late occurring phonemes in polysyllables were investigated. A study showed that lexical effects increase with word length (Pitt and Samuel, 2006) and the results obtained with other phoneme decision tasks also suggest that lexical effects are stronger for late occurring phonemes (see Introduction). So, we expected to obtain stronger lexical effect on late occurring phonemes in polysyllables in this experiment when compared to those obtained with monosyllables and bisyllables in Experiment 1.

We used six polysyllabic word-nonword contrasts (e.g. *hirondelle,* a word meaning *sparrow* vs. *hirontelle*, a nonword) in which the UP (e.g. nasal vowel "*on*") occurs before the phoneme for which identification responses were collected (e.g. /d/-/t/). Voicing continua were obtained by modifying the Voiceless Interval (VI), which corresponds to the sum of the closure silent interval and the VOT. The VI is a major cue for voicing perception in French medial stops (Saerens et al., 1989). Voicing responses in the word-nonword continua (e.g. *hirondelle – hirontelle*) were compared to control nonword--nonword continua (e.g. *ounondelle – ounontelle*). Let us note that the use of a nonword-nonword context as control for assessing the lexical effect is unusual in the study of lexical effects. This procedure was used here because it was not possible to find a word-nonword control in which the target phoneme was in the same phonetic context as the one prevailing in the contrast under study.

3.1 Experiment 2. Method

Stimuli. The speech material used for constructing the stimuli was a set of French words and nonwords pronounced by the same Belgian native speaker of French as in Experiment 1. There were 6 groups of utterances each made of one word and 3 related nonwords (see Table I). The mean syllabic rate was 3.18 per sec. Each word contained a single stop consonant located just after the UP and at the beginning of the third syllable. The stop was voiced for 3 words, namely for *hurluberlu*, *hirondelle* and *élongation* (meaning *crank*, *sparrow*, *strained muscle* respectively; the UP is underlined), and voiceless for the 3 other ones, namely *achoppement*, *lévitation* and *funiculaire* (meaning *stumbling block*, *levitation* and *funicular* respectively). The phonetic transcription of the stimuli in IPA symbols I given in Table II.

Lexical frequencies of the word stimuli are shown in Table I. The lexical neighborhood densities are also given in Table I, for each word and nonword stimulus.

The voicing continua were obtained by editing specific segments of the word and nonword utterances. The 24 (6 X 4) items were digitized at a sampling rate of 48 kHz with the help of the Soundtools software and stored on computer disk. The stimuli were again constructed with the Soundtools software. For each of the 6 series, 2 voicing continua were created, a word--nonword one (e.g. *hirondelle-hirontelle*) and a nonword-nonword one (e.g. *ounondelle-ounontelle*). The procedure was conceived in such a way as to minimize possible differences in the acoustic values of the voicing cues between the two continua. With the *hirondelle – hirontelle* contrast for example, the pitch periods present during the closure and the burst in *hirondelle* were progressively replaced by corresponding unvoiced segments extracted from *hirontelle*. The *onde-onte* segments were also inserted in the middle of *ounondelle* (a nonword) for creating an *ounondelle-ounontelle* continuum with the same acoustic cues as those used for the *hirondelle--hirontelle* one. While the first continuum is a word-nonword one, the second is nonword-nonword. The VOT and SI values of the continua endpoints are given in Table II. Ten voiceless segments were extracted from the original voiceless interval, corresponding to 100%, 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, and 10% of its total duration. For each segment, the SI and the VOT were in the same proportions, e.g. for the 80% voiceless interval, the SI amounted to 80% of the original SI and the VOT amounted to 80% of the original VOT. Each word-nonword and nonword-nonword continuum included 11 stimuli ranging from 0% SI (e.g. a stimulus including the unmodified *onde* segment) to 100% SI.

Participants. Twelve French-speaking students, 6 women and 6 men aged between 19 and 21 years took part in the experiment.

Procedure. The stimuli series were presented binaurally through head-phones. Listeners were tested individually. They were asked to identify the consonant located in the third syllable as belonging to the /b,d,g/ set or to the /p,t,k/ set by using one of two keys on a computer-controlled response box. Stimuli were delivered at a fixed rate of one per 5 seconds. An initial block of 50 stimuli was used for practice and responses to these trials were not included in the data. The 132 stimuli (12 series X 11 stimuli) were presented in an experimental series in which each stimulus appeared 10 times. The series were presented in two sessions.

3.2 Experiment 2. Results and discussion

Response functions for high and low expected rates of voiced responses, under the assumption that a lexical effect is present, are given in Figure 3.
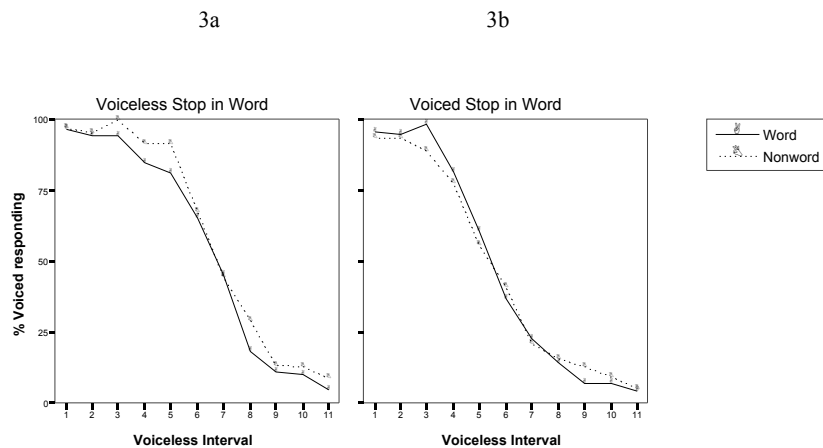

Figure 3: Results of Experiment 2. Mean percentages of voiced responses for the 6 word--nonword continua for which an advantage for voiceless responding was expected for words and corresponding nonword-nonword controls (left) and for the 6 ones for which an advantage for voiced responding was expected for words and corresponding nonword--nonword controls (right). Responses for word-nonword continua are indicated in continuous lines, those for nonword-nonword controls are indicated in broken lines.

High rates of voiced responding were expected for the three continua for which the word contained a voiced target phoneme (*hurluberlu-hurluperlu, hirondelle-hirontelle* and *élongation- éloncation)* when compared to the nonword-nonword controls (*olnunberlu-olnuperlu, ounondelle-ounontelle, ivongation-ivoncation*). Conversely, low rates of voiced responding were expected for the three other continua (*achoppement- achobement, lévitation-lévidation* and *funiculaire- funigulaire*) when compared to the nonword-

-nonword controls (*ourobement- ouropement, fusidation- fusitation, namigulaire- namiculaire*).

The results are presented in Figure 3. As expected, voiced scores are lower for the word-nonword continua with a *voiceless* stop in word compared to the nonword-nonword control continua (Figure 3a). For the word-nonword continua with a *voiced* stop in word compared to the nonword-nonword control continua, voiced scores are generally higher compared to the control continua although an inverse trend is also present at the upper-end of the continua (Figure 3b). As the word-nonword continua are compared to nonword-nonword continua, rather than to nonword-word continua as in Exp.1, the lexical effect corresponds to the sum of two differences in the present Experiment: the difference between the scores collected by the stimuli with a voiceless stop in word and those collected by the nonword-nonword control continua (between the two curves in Figure 3a) plus the difference between the scores collected by the stimuli with a voiced stop in word and those collected by the nonword-nonword control continua (between the two curves in Figure 3b). The lexical effect corresponded to the sum of these two differences in the statistical analyses.

The data were analyzed in a repeated measures ANOVA with Lexicality (voiced stop in word-voiceless in nonword vs. voiced in nonword- voiceless in word), and Context (/b,p/ context: hurluberlu/achoppement; /d,t/ context; hirondelle/lévitation; /g,k/ context: élongation/ funiculaire) as within--participants factors. The effect of Lexicality and the Lexicality X Context interaction were significant ($F_{(1,11)}$= 5.28, $p<.05$; $F_{(2,22)}$= 4.29, $p<.05$, respectively). Significance tests of Lexicality for each context are reported in Table I.
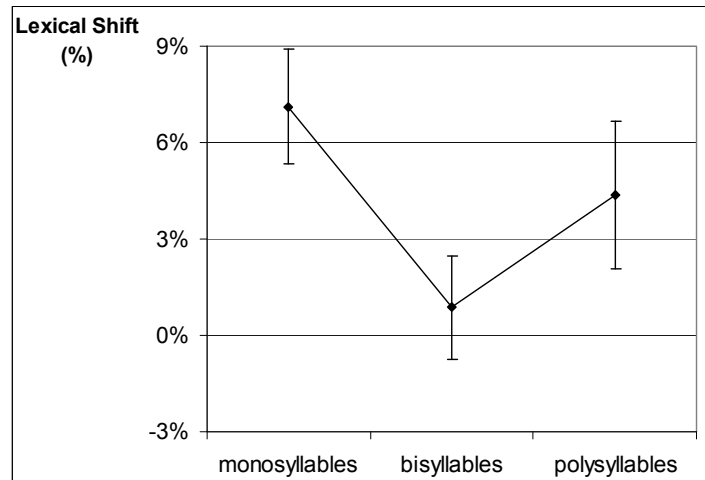

Figure 4: Means and SEs of the lexical shifts for monosyllables (Exp.1), bisyllables (Exp.1) and polysyllables (Exp.2).

As shown in Figure 4, the lexical effect in polysyllables was smaller than the one obtained for the monosyllables in Exp.2, a difference which was however not significant (t<1), and larger than the one obtained for the bisyllables in Exp.2, this difference being significant (t(30)= 3.08, p<.01). Remember that the difference in lexical effect between bisyllables and monosyllables was also significant (see Section 3.2).

The lexical effect in polysyllables was not significantly larger than the one obtained for monosyllables in Exp.1, contrary to what was expected from Pitt& Samuel (2006). Further, the effect of word length was non-monotonic: the lexical effect decreased from monosyllables to bisyllables and increased again from bisyllables to polysyllables.

4. General Discussion

The present experiments confirm the robustness of lexical effects in the identification of phonemes in monosyllables. Presence of these effects in French initial stops provides a cross-linguistic validation of those previously evidenced in English. The present data also reveal that the size of lexical effects depends on word length. Different factors known to interact with lexical effects, i.e. lexical frequencies, the number of real words embedded (Vroomen & de Gelder, 1997), stimulus duration and the location of the uniqueness point (Pitt & Samuel, 2006), covaried with word length in the stimuli used in this study. The number of embedded words and word duration increased with the number of syllables, the location of target phoneme was progressively delayed in words of increasing length and was located after the end of the word in monosyllables (Exp.1), just after the target or at the target phoneme in bisyllables (Exp.1) and before the target phoneme in polysyllables (Exp.2). However, these different factors contributed to amplify the effects of word length on the amount of lexical evidence and our purpose here was not to dissociate the effects of these factors but rather to examine the effects of the amount of lexical evidence on the size of the lexical shifts in phoneme identification.

Results for the different word lengths are summarized in Figure 4 using the procedure used by Pitt & Samuel (1993) for assessing the size of the lexical shift in different experiments. This procedure consists in calculating the mean identification change where the response curves shift away from one another. For monosyllables and for bisyllables (in Exp. 1), the lexical shift was assessed by averaging the differences in voiced responding between voiced and voiceless stops in words. For polysyllables (in Exp. 2), the lexical shift was assessed by summing up the differences in voiced responding either between voiced stops in words and those in nonwords or between voiceless stops in nonwords and those in words. The lexical shift was consistent and significant for initial phonemes in monosyllables (Exp.1), nearly absent for phonemes located at the beginning of the second syllable in bisyllables

(Exp.1) and present but relatively weak for internal phonemes located at the beginning of the third syllable in polysyllabic words (Exp. 2). It thus appears that the effect of lexical evidence is U-shaped: moderate increments of the amount of lexical evidence lead to a *decrease* in the size of the lexical effect and a larger increment of lexical evidence leads to an *increase* of the lexical effect.

Why should the effect of lexical evidence be U-shaped?

The U-shaped effect of word length on phoneme processing is difficult to explain by linear changes in lexical activation. The different variables which contribute or might possibly contribute to lexical activation, i.e. word frequency, density of lexical neighborhood, UP location… increase with word length in the present stimuli. As there is no reason why lexical activation should not steadily increase with word length in the present results, we can conclude that the U-shaped effect of lexical activation on phoneme processing is genuinely non-linear.

The non-linear effect of lexical activation on phoneme identification can be explained by the interplay between two different mechanisms. One of these mechanisms is driven by lexical evidence and corresponds to one of the various possible top-down interferences contemplated in the literature, depending on the model (see Introduction). "Lexical interferences" account for the increase in lexical effects as a function of word length, e.g. in polysyllables vs. bisyllables (in the present study) or in polysyllables vs. polysyllables (in Pitt & Samuel, 2006), but not for their decrease, e.g. in bisyllables vs. monosyllables (in the present study). Another type of mechanism is needed for explaining the drop in lexical effects with moderate increases in word length. Such a mechanism must be narrowly constrained by the amount of lexical evidence: it should be able to operate in the *absence* of lexical evidence and should not be able to operate with fairly large amounts of lexical evidence.

Suppose you hear a nonword such as "wiss". What can you do for trying to understand what has been said? It might be one of the various phonetically similar words: "wish", "was", "miss", "kiss", "will", etc… But, supposing the word was pronounced in isolation, you do not have any other clue for deciding and, besides guessing, the only solution is to re-analyze the acoustic input. Reanalyzing the input will *not* get you out of trouble if the speaker completely mispronounced one the phonemes. However, if one or several phonemes were ambiguous reanalyzing the input will give you reasonable chance to perceive the word intended by the speaker. Indeed, repeating the same ambiguous stimulus induces spontaneous changes in perception and this indicates that "…perceptual dynamics, not solely acoustic stimulus information, determines perception of speech categories" (Case, Tuller, Keslo & Ding, 1992; p.2413). Similar contrast effects were evidenced not only with several presentations of the same stimuli but also with a single adapter, both

in selective adaptation studies (Diehl, Elman & Buchwald McCusker, 1978) and in cross-modal adaptation (Bertelson, Vroomen, & de Gelder, 2003). These results suggest that re-analyzing the same stimulus when still in auditory memory can induce changes in phonetic identity in case of stimulus ambiguity and that these changes will generally generate word percepts, i.e. "wiss" will be changed in "wish", "was", "miss", "kiss", "will", etc… depending on which phonetic feature is ambiguous.

This example illustrates how reanalyzing the phonetic input might be useful for perceiving a word when both the phonetic and lexical information are ambiguous. We will refer to this kind of re-analysis as "*contrastive*" because it induces a change in phonetic identity similar to those evidenced by repetition of the same stimulus (Diehl et al. 1978; Bertelson et al., 2003). Further, we will use the term "*contrastive scanning*" because the re-analysis is supposed to be performed on all features present in the stimulus. One might postulate a more sophisticated process involving re-analysis of only the ambiguous features, but in the present state of the evidence this is unnecessary complex.

Contrastive scanning is able to operate in the absence of lexical evidence and it can therefore explain the presence of lexical effects in monosyllables. But is it also unable to process fairly large amounts of lexical evidence, a necessary condition to explain the drop of lexical effects in bisyllables? It might be useful to examine the role of auditory short-term memory to answer this question. Contrastive scanning necessitates the persistence of the acoustic stimulus in short term memory. Though durations of 0.5 to 1 sec. are reported for the duration of echoic memory, this represents the time taken for memory to fall to chance level and the half-time duration of echoic memory is probably much shorter (Huron & Parncutt, 1993). It is therefore possible that contrastive scanning can only operate on monosyllables (mean duration of 575 ms) because the decay in short term memory is too fast for retaining sufficiently detailed acoustic information on bisyllables (mean duration of 758 ms). Results from Pitt & Samuel (2006; Part 2) show that there is a sharp decrease in lexical effects as a function of reaction time in monosyllables vs. polysyllables, and that lexical effects in monosyllables are more sensitive to time compression than are those in polysyllables. This suggests that the mechanisms involved in both cases are different and that the mechanism involved in monosyllables is strongly constrained by time. According to Pitt & Samuel (2006), this might be due to "… dampering effect on lexical activation […] conditioned by word length" (p.1133). The dependency of contrastive scanning on auditory memory gives a specific content to the word length conditioned dampering postulated by these authors. And, going back to the interpretation of the present results, this would explain the drop in lexical effects for the bisyllables vs. the monosyllables.

Contrastive scanning would help recognize the right phoneme in situations where both the phonetic and the lexical information are ambiguous, such as

the one that prevails when the phonetic information is ambiguous in short words. Notice again that contrastive scanning is *not* driven by lexical evidence in support of a specific category and is immune from lexical influences on perceptual decisions. This mechanism is compatible with neuro-imagery data showing that lexical effects on phoneme recognition have a perceptual basis (Myers & Blumstein, 2008). However, although contrastive scanning is launched by a top-down signal, it is motivated by a lack of lexical evidence and the latter can thus not influence the outcome of the perceptual processing. Therefore, contrastive scanning is an entirely autonomous process and is compatible with models such as Race and Merge.

5. Conclusion

To sum up, two different mechanisms seem necessary to explain the U-shaped change in lexical effects as a function of word length evidenced in the present study. One of these mechanisms would be based on a re-scanning of the acoustic input. Contrastive scanning would operate in the absence of lexical evidence (in monosyllables) and it would be constrained by the persistence of the acoustic signal in short term memory. The other mechanism operates through top-down interferences and would only be effective with a fairly large amount of lexical evidence (in polysyllables). With moderate amounts of lexical evidence (in bisyllables), both mechanisms would be less effective, explaining the U-shaped evolution of lexical effects as a function word length. Further studies should enable to test this conjecture.

References

Bailly, G., & Guerti, M. (1991) Synthesis-by-rule for French. *Proceedings of the 12th International Congress of Phonetic Sciences*, p. 506-509. Aix-en-Provence: Université de Provence, Service des publications.

Burton, M.W., Baum, S.R., and 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.

Bertelson, P., Vroomen, J. & de Gelder, B. (2003) Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychological Science*, **14**, 592-597.

Case, P., Tuller, B., Keslo, J.A.S., & & Ding, M. (1992) Evidence for bistability in the region of perceptual boundaries. *Journal of the Acoustical Society of America*, **92**, 2413-2414.

Cole, R.A. (1973) Listening for mispronunciations: A measure of what we hear during speech. *Perception and Psychophysics*, **3**, 153-156.

Cole, R.A., Jakimik, J., & Cooper, W.E. (1978) Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America*, **64**, 44-56.

Connine, C., Titone, D., Deelman, T., & Blasko, D. (1997) Similarity mapping in spoken word recognition. *Journal of Memory and Language*, **37***, 463-480.*

Content, A., Mousty, P., & Radeau, M. (1990) Brulex: une base de données lexicales informatisées pour le français écrit et parlé. *L'Année Psychologique*, **90**, 551-566.

Cutler, A., & Norris, D. (1979) Monitoring sentence comprehension. In *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (W.E. Cooper & E.C.T. Walker, editors), pp. 113-134. Hillsdale, N.J.: Erlbaum.

Diehl, R., Elman, J.L. and Buchwald McCusker, S. (1978) Contrast effects on stop consonant identification. *Journal of Experimental Psychology, Human Perception and Performance*, **4**, 599-609.

Frauenfelder, U.H., Segui, J., & Dijkstra, T. (1990) Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 77-91.

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

Huron, D., & & Parncutt, R. (1993) An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology*, **12**, 154-171.

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

Luce, P.A. (1986) *Neighborhoods of words in the mental lexicon*. Unpublished Doctoral dissertation, Indiana University at Bloomington, Indiana.

Marslen-Wilson, W.D., & Welsh, A. (1978) Processing interaction and lexical access during word-recognition in continuous speech. *Cognitive Psychology*, **10**, 29-63.

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

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

McClelland, J., Mirman, D., & Holt, L. (2006). Are there interactive processes in

speech perception? *Trends in Cognitive Sciences,* **10***,* 363-369.

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 phoneme decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, **25**, 1363-1389.

McQueen, J.M., Norris, D., & Cutler, A. (2006) Are there really interactive processes in speech perception? *Trends in Cognitive Sciences*, **10**, 533.

Myers, E.B. & Blumstein, S.E. (2008) The neural bases of the lexical effect: An fMRI investigation. *Cerebral Cortex*, **18**, 278-288.

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., McQueen, J.M., & Cutler, A. (2000) Merging information in speech perception: Feedback is never necessary. *Behavioral and Brain Sciences*, **23**, 299--370.

Pitt, M.A. & Samuel, A.G. (1990) Attentional allocation during speech perception: How fine is the focus? *Journal of Memory and Language*, **29**, 611-632.

Pitt, M.A., & Samuel, A.G. (1993) An empirical and meta-analytical evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance*, **19**, 699-725.

Pitt, M.A., & Samuel, A.G. (1995) Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology*, **28**, 149-188.

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

Radeau, M., Morais, J., Mousty, Ph., & Bertelson, P. (2000) The effect of speaking rate on the role of the uniqueness point in spoken word recognition. *Journal of Memory and Language*, **42**, 406-422.

Repp, B.H. (1982) Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, **92**, 81-110.

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

Saerens, M., Serniclaes, W., & Beeckmans, R. (1989) Acoustic versus contextual factors in stop voicing perception in spontaneous French. *Language and Speech*, **32**, 291-314.

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

Scheffé, H. (1959) *The analysis of variance.* Wiley: New York.

Serniclaes, W. (1987) *Etude expérimentale de la perception du trait de voisement des occlusives du français*. Unpublished doctoral thesis. Université Libre de Bruxelles. http://lpp.psycho.univ-paris5.fr/person.php?name=WillyS

Serniclaes, W., & Wajskop, M. (1992) Phonetic versus acoustic account of feature interaction in speech perception. In *Analytic Approaches to Human Cognition, Proc. of the Conference in Honour of Paul Bertelson*. (J. Alegria, D. Holender, J. Junça de Morais, & M. Radeau, editors), pp. 77-91. Amsterdam: North-Holland.

Serniclaes, W., Beeckmans, R., & Radeau, M. (1995) Phonetic and lexical effects in speech perception. In *Levels in Speech Communication* (C.Sorin, J.Mariani, H. Meloni & J.Schoentgen, editors), pp. 39-51. Amsterdam: Elsevier.

Vitevitch, M.S., Stamer, M.K., & Sereno, J.A. (2008) Word length and lexical competition: Longer is the same as shorter. *Language and Speech,* **51**, 361-383.

Vroomen, J., & de Gelder, B. (1997) The activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **23**, 710-720.

Warren, R.M. (1970) Perceptual restoration of missing speech sounds. *Science*, **167**, 392-393.

Willy Serniclaes
Laboratoire de Psychologie de la Perception, CNRS & Université Paris Descartes.
Centre Biomédical des Saints Pères, 45, rue des Sts Pères, 75270, Paris cedex 06, France
willy.serniclaes@parisdescartes.fr

Renaud Beeckmans †
Unité de recherche en Neurosciences Cognitives, Université Libre de Bruxelles.
50, av. F.D. Roosevelt, 1050, Bruxelles, Belgium

Monique Radeau
Unité de recherche en Neurosciences Cognitives, Université Libre de Bruxelles
50, av. F.D. Roosevelt, 1050, Bruxelles, Belgium
moradeau@ulb.ac.be