Real-time lexical competitions during speech-in-speech comprehension

Véronique Boulenger, Michel Hoen, Emmanuel Ferragne, François Pellegrino, Fanny Meunier

Real-time lexical competitions during speech-in-speech comprehension

Véronique Boulenger [a], Michel Hoen [b], Emmanuel Ferragne [a], François Pellegrino [a], Fanny Meunier [a]

[a] *Laboratoire Dynamique du Langage, UMR 5596 CNRS – Université de Lyon, Institut des Sciences de l'Homme, 14 avenue Berthelot, 69363 Lyon Cedex 07, France*

[b] *Stem Cell and Brain Research Institute, INSERM U846 – Université Lyon 1, 18 avenue Doyen Lépine, 69675 Bron Cedex, France*

Corresponding author

Dr Véronique Boulenger

Laboratoire Dynamique du Langage – UMR 5596

Institut des Sciences de l'Homme

14 avenue Berthelot

69363 LYON Cedex (FRANCE)

Veronique.Boulenger@ish-lyon.cnrs.fr

Tel: +33(0)4.72.72.79.24 / Fax: +33(0)4.72.72.65.90

Abstract

This study aimed at characterizing the cognitive processes that come into play during speech-in-speech comprehension by examining lexical competitions between target speech and concurrent multi-talker babble. We investigated the effects of number of simultaneous talkers (2, 4, 6 or 8) and of the token frequency of the words that compose the babble (high or low) on lexical decision to target words. Results revealed a decrease in performance as measured by reaction times to targets with increasing number of concurrent talkers. Crucially, the frequency of words in the babble significantly affected performance: high-frequency babble interfered more strongly (by lengthening reaction times) with word recognition than low-frequency babble. This informational masking was particularly salient when only two talkers were present in the babble due to the availability of identifiable lexical items from the background. Our findings suggest that speech comprehension in multi-talker babble can trigger competitions at the lexical level between target and background. They further highlight the importance of investigating speech-in-speech comprehension situations as they may provide crucial information on interactive and competitive mechanisms that occur in real-time during word recognition.

Key words: speech-in-noise; informational masking; lexical competition

1. Introduction

Under ecological conditions, speech is hardly ever perceived in ideal acoustic conditions, be it in the chaos of a traffic-jam or in a babbling crowd. Still our cognitive system is able to compensate for such degradation allowing us to understand the delivered message. Speech comprehension in noisy environments also appears to be the primary problem experienced by hearing-impaired people, strongly affecting their communication abilities. Understanding the processes underlying speech-in-noise comprehension, and particularly comprehension of speech in speech, therefore constitutes a challenge that scientists have to face given its major social impact. In the present study, we aimed at characterizing the cognitive processes that come into play during speech-in-speech comprehension by examining lexical competitions during word recognition in cocktail party situations (Cherry, 1953).

Most psycholinguistic models of language processing postulate that speech recognition is supported by an interactive system in which bottom-up processes (going from low-level acoustic information to higher-level information such as meaning) are combined with top-down processes where high-level information may modulate lower-level processing (*e.g.*, Davis and Johnsrude, 2007). If the representation of a word becomes active, this lexical activation might increase the activation of the phonemes that compose the word, allowing them to be recognized with less acoustic evidence than they would need without the top-down lexical influence (Samuel, 2001). One example illustrating such lexical feedback is the so-called "word advantage", namely the fact that speech sounds in words are recognized more quickly than speech sounds in non-words (Mirman et al., 2005; Rubin et al., 1976; Samuel, 1997, 2001). Psycholinguistic models, although making different proposals regarding the nature of the competitors, further assume that word recognition results from strong competitive mechanisms between simultaneously activated lexical candidates (Marslen-

Wilson et al., 1996; McClelland and Elman, 1986). Identifying the processes at play during speech-in-speech comprehension may thus provide crucial information on interactions between the different levels of speech processing and on competition between these levels.

When listening to speech in noise, two types of masking can be distinguished (Bronkhorst, 2000; Brungart, 2001): *energetic masking* refers to the overlap in time and frequency between target speech and background noise so that portions of the target signal are rendered inaudible or at least unintelligible. Higher-level *informational masking* occurs when target and masker signals are both audible but the listener is unable to disentangle them. Such masking may relate to auditory scene analysis issues such as an inability to parse the acoustic scene into distinct messages (Alain et al., 2001, 2005), but it also results from an inability to successfully attend to target speech while inhibiting attention to irrelevant auditory objects. In the context of speech-in-speech comprehension, energetic masking has been shown to some extent to play a relatively small role in the overall masking phenomenon (Brungart et al., 2006). In contrast, informational masking becomes highly relevant as multi-talker babble carries linguistic information (phonetic, lexical and semantic) that may compete with the processing of target speech. When target and masker are speech, both may indeed elicit activity in the language system leading to interference effects at the cognitive level. This masking effect is more likely to emerge when only a few simultaneous talkers are present in the babble as speech-specific information is still available from the background (Brungart, 2001; Hoen et al., 2007; Simpson and Cooke, 2005; Van Engen and Bradlow, 2007). Besides, informational masking may occur along the different psycholinguistic dimensions that characterize speech sounds, namely pitch, phonemic and lexical information, etc. For instance, previous studies reported that differences in the vocal characteristics of the competing talkers, such as gender differences between target and masker, can considerably improve the intelligibility of target speech (Brungart, 2001b; Brungart et al., 2001; Festen and Plomp, 1990). In a recent study

(Hoen et al., 2007), we further examined the differential effects of acoustic-phonetic and lexical content of natural speech babble or time-reversed speech babble (which contains only partial phonetic information) on target word identification. The results reveal that in the presence of 4 concurrent talkers, identification rates are poorer in the natural than in the reversed babble. We interpreted this finding as evidence that 4-talker natural babble causes increased informational masking that may stem from increased lexical competition effects triggered by the availability of identifiable lexical items from background.

The present study aimed at further breaking down informational masking into its different constituents by examining real-time *lexical* competitions between target and background in speech-in-speech comprehension situations. Van Engen and Bradlow (2007) recently showed that the language used as masker can affect the intelligibility of target speech: sentence recognition is poorer when the background babble is the same language as the one of the target speech compared to when they are different (see also Garcia Lecumberri and Cooke, 2006 and Rhebergen et al., 2005 for similar findings). Although these results provide evidence for informational masking in the form of linguistic interference, the exact linguistic features that contribute to this effect are still unspecified. Such linguistic interference may in fact occur at the lexical (i.e. whole-word effect), sublexical (*e.g.* differences in phoneme inventories and syllable structures of the languages) or even at the prosodic level. These different levels of interference may further imply different mechanisms of compensation of the degraded message that still need to be investigated.

Here we specifically addressed the contribution of lexical factors to speech-in-speech perception. We sought to determine to what extent lexical information from background babble is processed, even not consciously, by the listener and can therefore interfere with target speech recognition. To this aim, we examined whether and how the frequency of words that compose the babble influences lexical access to target speech. Frequency has been widely

shown to affect word recognition: high-frequency words are better recognized than low-frequency words both in terms of reaction times and error rates (Connine et al., 1990; Monsell, 1991; Taft and Hambly, 1986). Most current models of spoken word recognition, although not considering the particular case of degraded speech perception, assume that frequency affects the activation levels of competing lexical candidates during lexical access. High-frequency words would be processed faster than low-frequency words because frequency determines either the baseline activation level of each lexical unit (McClelland and Rumelhart, 1981; Marslen-Wilson, 1990) or the strength of the connections from sublexical to lexical units (MacKay, 1982, 1987). In distributed learning models, the representations of high-frequency words would be activated more rapidly because high-frequency mappings are better learned, resulting in stronger connection weights (Gaskell and Marslen-Wilson, 1997; Plaut et al., 1996). In our study, we hypothesized that if listeners are sensitive to lexical factors in the background babble, lexical competition between target and babble should vary depending on the frequency of words that compose this babble. More precisely, we predicted that high-frequency words in the babble should hinder more strongly target word recognition than low-frequency words due to increased competitions within the mental lexicon. To test this prediction, we asked healthy participants to perform a lexical decision task on target items in the presence of competing babble in which the number of simultaneous talkers (2, 4, 6 or 8) and the frequency of words (high or low) were manipulated. The lexical decision task was used as it is considered a reliable index of lexical access in the brain. Investigating whether speech comprehension is affected by lexical factors from the babble indeed requires using an on-line task that measures access to the mental lexicon rather than a task that involves post-lexical processes taking place after lexical activation to target words occurs.

2. Materials and methods

2.1. Participants

Thirty-two healthy volunteers, aged 18-26 years, participated in the experiment. All were French native speakers and right-handed with no known hearing or language disorders. They signed a consent form and were paid for their participation.

2.2. Stimuli

2.2.1. Multi-talker babble

Two lists of 1250 words each were created from the French lexical database Lexique 2 (New et al., 2004). The first list F+ included words of high frequency of occurrence ($45.03 < F+ < 13896.7$ per million) whereas the second list F- included low-frequency words ($0.03 < F- < 1$ p/m). The words in both lists were matched for length in letters and number of syllables. Eight French native speakers (4 females, 4 males) recorded the two lists in a sound-attenuated room (sampling rate 44 kHz, 16 bits). The order of words in the lists was randomized and different for each talker. These 16 recordings (8 talkers x 2 word frequency) gave the babble signals. Individual recordings were checked and modified according to the following protocol: (i) removal of silences and pauses of more than 1s, (ii) suppression of words containing pronunciation errors, (iii) noise reduction optimized for speech signals, (iv) intensity calibration in dB-A and normalization of each source at 70 dB-A. Each recording was then segmented into 30 chunks of 4 s for which two acoustic parameters (pitch and speech rate) were estimated to control for their potential influence. Individual pitch was computed as the median value of *F0* within each chunk, and averaged over the 30 excerpts using Praat (Boersma & Weenink, 2009). Individual speech rate was estimated as the average syllabic rate over the 30 chunks, with an automatic vowel detection algorithm (Pellegrino & André-Obrecht, 2000; Pellegrino et al., 2004).

The analysis of these acoustic parameters shows obvious individual differences among the 8 talkers and slight differences between the F+ and F- conditions (Table 1). The overall average speech rate is 5.18 and 5.16 syllables per second for female and male talkers respectively. The average pitch is 142 Hz and 113 Hz for female and male talkers respectively. No systematic tendency between F+ and F- can be observed in terms of pitch: one half of the talkers (1, 2, 4 and 7) exhibits a lower pitch for the F+ condition while the other shows the opposite.

		Talker							
Lexical Frequency	Acoustic features	1 (F)	2 (M)	3 (M)	4 (F)	5 (M)	6 (F)	7 (F)	8 (M)
F-	Average *F0* (Hz)	147.64	114.71	108.80	141.94	122.67	140.78	138.78	105.86
	F0 SD (Hz)	1.42	1.53	1.90	1.53	2.28	1.37	1.27	1.34
	Average Speech Rate (syl/s)	5.78	4.94	5.33	5.27	6.07	5.10	5.37	5.35
	Speech Rate SD (syl/s)	0.47	0.55	0.61	0.75	0.54	0.48	0.44	0.54
F+	Average *F0* (Hz)	145.42	110.00	109.23	139.95	123.32	147.04	137.09	107.44
	F0 SD (Hz)	2.20	1.93	1.50	1.72	1.92	1.24	1.04	2.05
	Average Speech Rate (syl/s)	4.88	4.13	5.09	5.06	5.33	4.73	5.25	5.04
	Speech Rate SD (syl/s)	0.72	0.77	0.57	0.49	0.58	0.44	0.67	0.61

Table 1. Average pitch (Hz) and mean speech rate (number of syllables/s) together with standard deviations (SD) for each talker included in the babble signals. The acoustic features were estimated for the F+ and F- conditions separately. Each talker sex (M: Male; F: Female) is given along with their identification (1 to 8).

Babble signals were generated from the individual chunks with 2, 4, 6 and 8 talkers (half female, half male) and for both F+ and F- conditions, resulting in 8 cocktail-party sounds (4 talker numbers x 2 frequencies of words in the babble): T2F+, T2F-, T4F+, T4F-, T6F+, T6F, T8F+ and T8F-. Since slight acoustic variation was revealed by the preliminary analysis (Table 1), talkers 3 and 4, involved in the 2-talker babble, were carefully selected to avoid any acoustic bias between the F+ and F- conditions. This is particularly important as this babble constitutes the critical condition where interference between target speech and babble is more likely to emerge. More precisely, for both talkers, 30 F+ chunks were matched with 30 F- chunks based on minimal distance in terms of *F0* and speech rate. A two-factor multivariate

analysis of variance, with *F0* and speech rate as dependent variables, confirmed a significant difference between talkers as revealed by Wilks' Lambda ($F(2,155) = 16814.72$, $p < .001$) and a non-significant difference between F+ and F- conditions ($F(2,155) < 1$, ns). The interaction was also non-significant ($F(2,155) < 1$, ns).

2.2.2. Target words and pseudo-words

One hundred and twenty mono-syllabic, tri-phonemic French words were selected from Lexique 2 (New et al., 2004) in a middle range of frequency of occurrence (mean = 53.89 p/m, SD = 69.77). All words were CVC, CCV, CYV or CVY words. One hundred and twenty mono-syllabic pseudo-words were also constructed by changing one phoneme in target words (*e.g.* pseudo-word "rambe" / word "jambe"). All pseudo-words respected the phonotactic rules of French. The stimuli were recorded in a sound-attenuated booth by a female French native speaker different from the talkers who recorded the babble. The target words and pseudo-words were hyper-articulated as they were produced in isolation.

2.2.3. Stimuli and word lists

The stimuli consisted of 120 target words and 120 target pseudo-words mixed with 4s chunks of multi-talker babble at a signal-to-noise ratio (SNR) set to zero. Target items were inserted 2.5 s from the start of the stimulus so that participants always had the same exposure to the babble before target speech was presented. Individual babble mixed with target files were further normalized at an equivalent intensity of 70 dB-A. As this resulted in some modulation of the intensity of the final multi-talker babble sounds, a final randomized intensity roving over a ±3 dB range in 1 dB steps was applied to each stimulus.

Sixteen different experimental lists (8 for words and 8 for pseudo-words) – the same list being seen by 4 participants – were generated. Each list contained every target item only once

to avoid repetition effects (i.e. each participant heard each target word and each pseudo-word once). In the end, each list was made up of 240 stimuli (120 words and 120 pseudo-words), with 15 words and 15 pseudo-words being presented in each of the 8 babble conditions (e.g. 15 words in T2F-, 15 pseudo-words in T2F-, 15 words in T2F+, 15 pseudo-words in T2F+ etc.). Across lists, all target words were presented against all babble conditions. Within each list, the order of words and pseudo-words and of babble conditions was randomized across participants.

2.3. Procedure

Participants were comfortably seated in a quiet room facing a computer monitor. Stimuli were presented diotically over headphones at a comfortable sound level. Participants were instructed to attentively listen to the stimuli and to perform a lexical decision task on target items that were presented in background babble. They had to decide as quickly and accurately as possible whether the target was a word or not by pressing one of two pre-selected keys on a computer keyboard. For half of the participants, response to words was given with the right hand and response to pseudo-words with the left hand. The reverse was true for the other half. The task was self-paced, that is, participants pressed the space bar on the keyboard to move from trial to trial. They could listen to each stimulus no more than once. Stimuli were presented in a randomized order by means of E-Prime software (Psychology Software Tools, Inc., Pittsburgh, PA). Before the testing phase, participants were given 16 practice items (8 words and 8 pseudo-words, different from the stimuli from the experimental lists) to accommodate themselves to stimulus presentation mode and target voice. To ensure that they paid attention to the task, and that they did not respond to words by chance, participants were informed that they could occasionally be asked to transcribe the word they had just heard on a piece of paper. As this was a control task irrelevant for the purpose of this study, we did not

analyze these results. However, mean percentage of correct responses was 70.1 % (excluding responses which were phonologically similar to the target; *e.g.* "lime" instead of "rime") suggesting that participants were indeed attentive to target speech and therefore engaged on the task.

3. Results

3.1. Behavioural results

Reaction times (RTs: time-interval between the onset of the target stimulus and the button press; in milliseconds) for target word identification were measured. Trials for which participants gave no response or made mistakes (word response for a pseudo-word or vice-versa) were considered as errors and were not included in the analysis. A two-way repeated measures analysis of variance (ANOVA) was conducted considering RT as the dependent variable and Talker number (2 vs. 4 vs. 6 vs. 8) and Frequency of words in the babble (F+ vs. F-) as within-subjects factors.

Results revealed a significant main effect of talker number on participants' RTs for target word identification ($F(3, 93) = 3.77$, $p = .01$). Mean RTs were faster when 2 talkers were present in the babble (1017 ms, SD = 84) compared to the 4- (1039 ms, SD = 100), 6- (1045 ms, SD = 94) and 8-talker conditions (1046 ms, SD = 94; Fig. 1). These latter three conditions (T4, T6 and T8) did not significantly differ from each other.

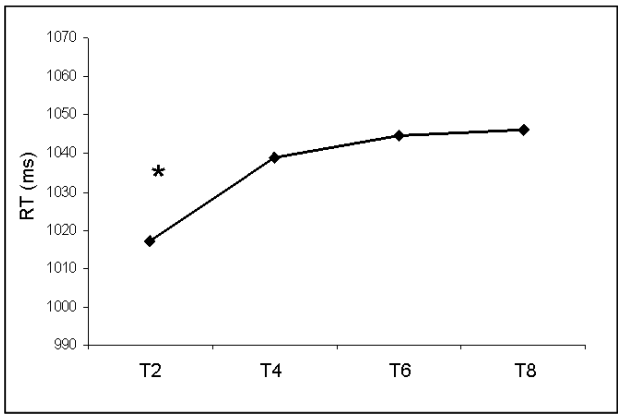

Fig. 1. Mean reaction times (ms) for target word identification when 2- (T2), 4- (T4), 6- (T6) and 8-talkers (T8) were present in the babble. * indicates a significant difference between T2 and other conditions.

Importantly, a significant main effect of frequency of words in the babble also emerged ($F(1, 31) = 11.58$, $p = .001$), with longer mean RTs when words in the babble were highly frequent (1045 ms, SD = 89) compared to when they were less frequent (1028 ms, SD = 97; Fig. 2). A significant interaction was also observed between the two factors ($F(3, 93) = 2.82$, $p = .04$), indicating that the effect of babble word frequency depended on the number of talkers present in the babble. Post-hoc comparisons (HSD Tukey) showed that words embedded in high-frequency babble were responded to later than words embedded in low-frequency babble only in the 2- ($p = .03$) and 8-talker conditions ($p = .03$; Fig. 3).

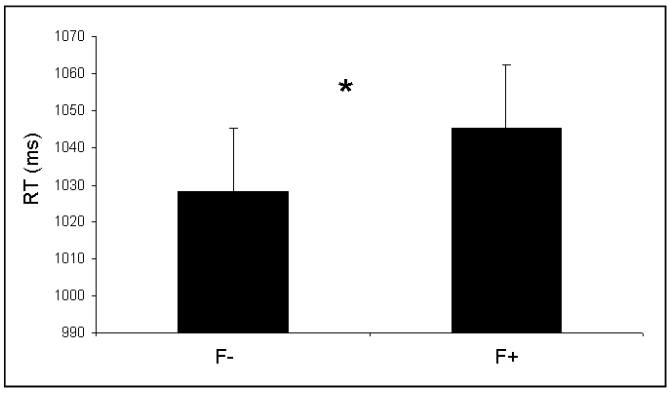

Fig. 2. Mean reaction times (ms) for target word identification when babble was composed of low-frequency words (F-) and high-frequency words (F+). Standard errors are reported. * indicates a significant difference between the conditions.

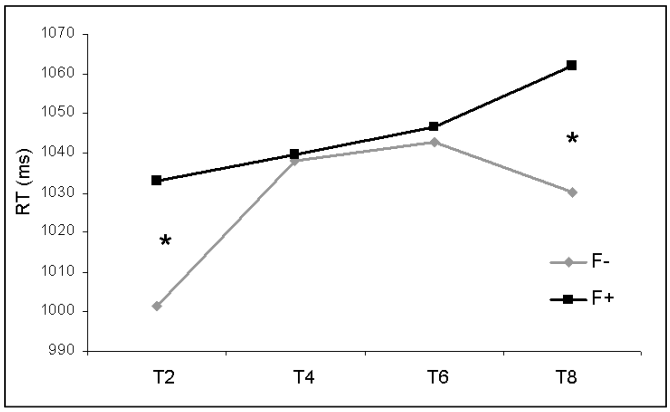

Fig. 3. Direct comparison of the effect of frequency of words in the babble (F- vs. F+) on mean reaction times (ms) to target words in each of the multi-talker conditions. Note the significant difference (*) between F- and F+ in the 2- and 8-talker conditions.

3.2. Preliminary discussion of the results

Lexical decision to target words masked by competing babble was influenced by the number of talkers as well as by the frequency of words that compose this babble. This latter effect was significant only in the 2- and 8-talker conditions, with high-frequency babble lengthening reaction times compared to low-frequency babble. In the 2-talker babble, the competing speech signal is still intelligible: besides energetic masking, the lexical content of the babble and particularly the frequency of words that compose this babble can therefore trigger lexical competitions and affect target word recognition. The effect of babble word frequency in the 8-talker condition is by contrast unlikely to result from such lexical informational masking. In fact, the large number of interfering talkers in this babble causes increased spectro-temporal saturation such that complete lexical items may no longer be available. This babble may instead act as both an energetic and an informational masker at a lower linguistic level (acoustic-phonetic) than the 2-talker babble (Hawley et al., 2004; Hoen

et al., 2007). To test this hypothesis, we performed an acoustic analysis following the method developed by Hoen et al. (2007) which aimed at evaluating the effect of spectro-temporal saturation on speech comprehension.

3.3. Acoustic analysis of multi-talker babble

The acoustic analysis was performed on the 240 (30 chunks x 4 different numbers of talkers x 2 word frequencies) babble signals that were used in the experiment. Each chunk was first segmented in sub-phonemic components using a statistical algorithm of detection of acoustic changes (André-Obrecht, 1988). Hoen et al. (2007) indicated that mean cepstral variation (i.e. the cepstral distance between two consecutive segments) is maximal for 1-talker speech and decreases monotonically when the number of considered speakers increases, resulting in smoothing over segmental differences (i.e. increased saturation). Following Hoen et al. (2007), mean cepstral variation was extracted for each chunk. A two-way ANOVA including talker number and babble word frequency as factors and mean cepstral variation as dependent variable was conducted. Results first revealed that an increase in the number of simultaneous talkers monotonically increased saturation as measured by mean cepstral variation ($F (3, 87) = 971.92$, $p < .001$). The frequency of words in the babble also significantly affected this parameter ($F (3, 29) = 19.41$, $p < .001$), with cepstral variation being larger in the F+ than in the F- babble. The interaction between talker number and babble word frequency was not significant ($F (3, 87) < 1$, ns). Post-hoc analysis (HSD Tukey) however revealed that mean cepstral variation was significantly larger in the F+ than in the F- babble only in the 8-talker condition ($p = .03$; Fig. 4). In agreement with our hypothesis, when 8 talkers were present in the babble, acoustic/energetic features distinguished between high-frequency and low-frequency word babble.

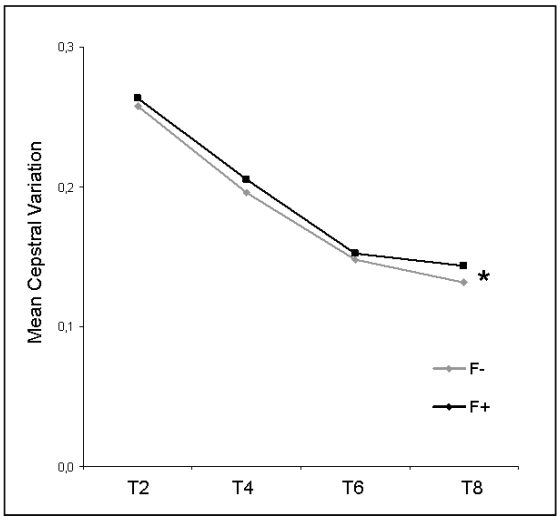

Fig. 4. Mean cepstral variation among segments in the 2- (T2), 4- (T4), 6- (T6) and 8 (T8) multi-talker babble as a function of frequency of words within these babble (high-frequency F+ vs. low-frequency F-). Note the significant difference (*) between the two

4. General discussion

This study investigated lexical competitions during speech-in-speech comprehension using a lexical decision task. We were particularly interested in determining to what extent lexical information from multi-talker babble is processed and can compete with speech recognition. Behavioural results first showed better recognition of target words in terms of reaction times when only 2 talkers were present in the babble. Increasing the number of talkers to 4, 6 and 8 decreased performance in the same way. This result is in line with previous studies showing that a larger number of interfering talkers reduces speech intelligibility due to progressively increased spectro-temporal saturation (Brungart, 2001; Hoen et al., 2007; Simpson and Cooke, 2005). When only a few talkers are present in the babble, masking effects may actually be more easily overcome because of clear acoustical distinctions between voices (Brungart, 2001) or because listeners can rely on asynchronies in the dynamic variations of the concurrent streams that cause transient gaps in the babble during which they can listen to target signals (Hoen et al., 2007). With an increasing number of talkers however, the dynamic modulations from the additive sources are progressively averaged, thus decreasing the temporal gaps free for listening to target words (Bronkhorst and Plomp, 1992; Drullman and

Bronkhorst, 2000). This phenomenon has been considered as informational masking occurring at the acoustic–phonetic level (Hoen et al., 2007).

Crucially, the study also revealed that the frequency of words in the babble can significantly affect performance, with high-frequency words being more detrimental to target word recognition than low-frequency words, but only in the 2- and 8-talker babble. The lexical effect we observed in the 2-talker babble is in agreement with the study by Van Engen and Bradlow (2007) who also found linguistic interference between babble of different languages and target speech when only 2 talkers were present in the background. In such babble, the interfering speech signal is still intelligible, high-level (lexical and semantic) information is therefore more likely to affect the ability of the listener to single out and understand target speech (Hawley et al., 2004). Our findings show that this lexical interference is stronger when words in the babble are highly frequent, as these words may strongly activate the mental lexicon, hence increasing lexical competition between target speech and background. As mentioned in the introduction, psycholinguistic models of language processing postulate that word recognition is the result of strong competitive mechanisms between concurrently activated lexical candidates (Davis and Johnsrude, 2007; Marslen-Wilson et al., 1996; McClelland and Elman, 1986). Since less information is required to activate highly frequent words compared to low frequent words, in a particularly challenging situation, these words may be stronger competitors during word recognition than low-frequency words. Accordingly, when listening to speech in high-frequency babble, the competitions may be maximal, "overloading" the language processing system: lexical access to target words may be more difficult, resulting in worse performance for target word recognition (lengthened reaction times). As far as informational masking is concerned, this lexical interference may relate to top-down control of selective attention allowing the listener to select a target amongst distracters as well as to inhibit attention to irrelevant auditory

information (Alain et al., 2001, 2005; Alain & Izenberg, 2003; Bregman, 1990). Alternatively, this effect could be related to earlier bottom-up processes involved in auditory scene analysis that are independent of attention, such as acoustic differences between high- and low-frequency babble. However, the matching of high- and low-frequency words in the babble in terms of average pitch and speech rate, together with the results of the acoustic analysis that revealed no difference in mean cepstral variation between the two types of 2-talker babble (F- and F+), rule out such an explanation. On the contrary, the effect of babble word frequency in the 8-talker babble seems to result from such acoustic differences between high- and low-frequency conditions, as revealed by the acoustic analysis, and not from lexical informational masking.

In the frame of psycholinguistic models of language processing, our results suggest that degraded speech comprehension, which requires restoration or compensation of partial linguistic information, constitutes a privileged situation in which to look at the importance of bottom-up and top-down interactions during speech processing (see also Davis and Johnsrude, 2007). Particularly, speech-in-speech comprehension offers a natural example of speech perception where competitions between different candidates at various linguistic levels (phonological, lexical and semantic) can be measured on-line during word identification. Using speech-in-speech comprehension paradigms may thus improve theoretical models of lexical access in speech comprehension by providing experimental evidence of an otherwise only indirectly approached issue: linguistic information competition phenomena.

Cocktail party situations may further allow testing the issue of subliminal processing in multi-talker babble as participants can not consciously process all the overlapping speech streams. Whereas subliminal processing has been extensively studied in visual word processing (Forster and Davis, 1984), it has hardly ever been tested in the auditory modality

due to the lack of a suitable paradigm. Previous attempts to develop auditory masking techniques, either by attenuating word stimuli embedded in white noise (Moore, 1995) or by presenting speech in parallel using dichotic listening procedures and manipulating attention to prevent awareness (Dupoux et al., 2002), have failed to produce any satisfactory evidence of any kind of activation in the case of unawareness of the prime. Recently however, Kouider and Dupoux (2005) reported repetition priming effects when auditory prime stimuli were time-compressed and hidden within a stream of spectrally similar unintelligible speech-like noise. In this situation, the prime is presented in a continuous stream so that it does not show up as a discrete acoustic event and is therefore not noticed by participants. The authors interpreted their findings as evidence that subliminal priming for spoken word processing is feasible and that lexical processing occurs at an early and unconscious stage of speech perception. Interestingly enough, speech-in-speech comprehension paradigms such as the one we used in the present study offer a more natural situation where auditory subliminal priming can be tested. Words that compose the multi-talker babble are indeed not perceptually nor physically degraded, they are surrounded by other speech streams so that they are not perceived as discrete units and the temporal relationship (stimulus onset asynchrony) between the prime and the target can be easily manipulated (the prime can precede the target immediately or not, or it can overlap in time the target). By varying prime-target relation (phonologically, lexically, morphologically or semantically), speech-in-speech paradigms could be used as a new tool to investigate the processing levels at which subliminal speech priming occurs.

5. Conclusion

The present study revealed that speech comprehension in multi-talker babble triggers competitions between target and background at the lexical level. Lexical factors such as the

frequency of words that compose the babble contribute to informational masking during speech-in-speech recognition: babble made of high-frequency words has a stronger adverse effect on speech recognition than low-frequency word babble. These lexical competitions are particularly salient when only a few simultaneous talkers are present in the babble due to the availability of identifiable lexical items from the background. Such findings highlight the importance of examining speech-in-speech comprehension situations as it could serve as a new paradigm to study interactive and competitive mechanisms that occur in real-time during language processing.

Acknowledgements

This project was carried out with financial support from the European Research Council (SpiN project to Fanny Meunier). We would like to thank Claire Grataloup for allowing us to use the materials from her PhD. We would also like to thank the anonymous Reviewer and the Editor for their very helpful comments.

References

Alain, C., Arnott, S.R., Picton, T.W., 2001. Bottom-up and top-down influences on auditory scene analysis: evidence from event-related potentials. J Exp Psychol: Hum Percept Perform, 27(5), 1072-1089.

Alain, C., Izenberg, A., 2003. Effects of attentional load on auditory scene analysis. J Cogn Neurosci, 15(7), 1063-1073.

Alain, C., Reinke, K., He, Y., Wang, C., Lobaugh, N., 2005. Hearing two things at once: neurophysiological indices of speech segregation and identification. J Cogn Neurosci, 17(5), 811-818.

André-Obrecht, R. (1988). A new statistical approach for the automatic segmentation of

continuous speech signals. IEEE Trans. on ASSP, 36, 29-40.

Bregman, A.S., 1990. Auditory scene analysis: The perceptual organization of sounds. London, MIT Press.

Bronkhorst, A., 2000. The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. Acustica 86, 117–128.

Bronkhorst, A., Plomp, R., 1992. Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. J. Acoust. Soc. Am. 92, 3132–3138.

Brungart, D.S., 2001. Informational and energetic masking effects in the perception of two simultaneous talkers. J. Acoust. Soc. Am. 109, 1101–1109.

Brungart, D.S., Simpson, B.D., Ericson, M.A., Scott, K.R., 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J. Acoust. Soc. Am. 110, 2527–2538.

Brungart, D.S., Chang, P.S., Simpson, B.D., Wang, D., 2006. Isolating the energetic component of speech-on-speech masking with ideal time frequency segregation. J. Acoust. Soc. Am. 120, 4007–4018.

Cherry, E., 1953. Some experiments on the recognition of speech, with one and two ears. J. Acoust. Soc. Am. 25, 975–979.

Connine, C.M., Mullenix, J., Shernoff, E., Yelen, J., 1990. Word familiarity and frequency in visual and auditory word recognition. J. Exp. Psychol.: Learn. Mem. Cogn. 16(6), 1084–1096.

Davis, M.H., Johnsrude, I.S., 2007. Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hear. Res. 229(1-2), 132–147.

Dupoux, E., Kouider, S., Mehler, J., 2002. Lexical access without attention? Exploration using dichotic priming. J. Exp. Psychol.: Hum. Percep. Perfor. 29, 172–183.

Drullman, R., Bronkhorst, A., 2000. Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation, J. Acoust. Soc. Am. 107, 2224–2235.

Festen, J., Plomp, R., 1990. Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing. J. Acoust. Soc. Amer. 88, 1725–1736.

Forster, K.I., Davis, C., 1984. Repetition priming and frequency attenuation in lexical access. J. Exp.Psychol.: Learn. Mem. Cogn. 10, 680–698.

Garcia Lecumberri M.L., Cooke, M., 2006. Effect of masker type on native and non-native consonant perception in noise. J. Acoust. Soc. Am. 119, 2445–2454.

Gaskell, M.G., Marslen-Wilson, W.D., 1997. Discriminating local and distributed models of competition in spoken word recognition. In: Shafto, M.G., Langley, P. (Eds.), Proc. Nineteenth Annual Conf. of the Cognitive Science Society, pp. 247–252.

Hawley, M.L., Litovsky, R.Y., Culling, J.F., 2004. The benefit of binaural hearing in a cocktail party: effect of location and type of interferer", J. Acoust. Soc. Am. 115(2), 833–843.

Hoen, M., Meunier, F., Grataloup, C., Pellegrino, F., Grimaut, N., Perrin, F., Collet, L., 2007. Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. Speech Communication, 49, 905–916.

Kouider, S., Dupoux, E., 2005. Subliminal speech priming. Psychol. Sci. 16(8), 617–625.

MacKay, D.G., 1982. The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. Psychol. Rev. 89, 483–506.

MacKay, D.G., 1987. The organization of perception and action: A theory for language and other cognitive skills, Springer-Verlag, New York.

Marslen-Wilson, W.D., 1990. Activation, competition, and frequency in lexical access. In: Altmann, G. (Ed.), Cognitive Models of Speech Perception: Psycholinguistic and Computational Perspectives. MIT Press, Cambridge, USA, pp. 148–172.

Marslen-Wilson, W.D., Moss, H.E., van Halen, S., 1996. Perceptual distance and competition in lexical access. J. Exp. Psychol.: Hum. Percept. Perform. 22, 1376–1392.

McClelland, J.L., Rumelhart, D.E., 1981. An interactive activation model of context effects in letter perception: Part 1. An account of Basic Findings. Psychol. Rev. 88, 375–407.

McClelland, J.L., Elman, J.L., 1986. The TRACE model of speech perception. Cogn. Psychol. 8, 1–86.

Mirman, D., McClelland, J.L., Holt L.L., 2005. Computational and behavioral investigations of lexically induced delays in phoneme recognition. J. Mem. Lang. 52(3), 424–443.

Monsell, S., 1991. The nature and locus of word frequency effects in reading. In Besner, D., Humphreys, G.W. (Eds.), Basic processes in reading: Visual word recognition, pp. 148–197, Hillsdale: Lawrence Erlbaum Associates.

Moore, T.E., 1995. Subliminal self-help auditory tapes: An empirical test of perceptual consequences. Can. J. Behav. Sci. 27, 9–20.

New, B., Pallier, C., Brysbaert, M., Ferrand, L., 2004. Lexique 2: A New French Lexical Database. Behav. Res. Methods Instrum. Comput. 36(3), 516–24.

Pellegrino, F. & André-Obrecht, R., 2000. Automatic language identification: an alternative approach to phonetic modeling. Signal Processing, Special Issue on Emerging Techniques for Communication Terminals, 80, 1231-1244.

Pellegrino, F., Farinas, J., Rouas, J.L, 2004. Automatic Estimation of Speaking Rate in Multilingual Spontaneous Speech. In : International Conference on Speech Prosody 2004, Nara, Japan. B. Bel, I. Marlien (Eds.), ISCA Special Interest Group on Speech Prosody (SproSIG), ISBN 2-9518233-1-2, p. 517-520.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S., Patterson, K.E., 1996. Understanding normal and impaired word reading: computational principles in quasi-regular domains. Psychol. Rev. 103, 56–115.

Rhebergen, K.S., Versfeld, N.J., Dreschler, W.A., 2005. Release from informational masking by time reversal of native and non-native interfering speech (L). J. Acoust. Soc. Am. 118, 1274–1277.

Rubin, P., Turvey, M.T., Van Gelder, P., 1976. Initial phonemes are detected faster in spoken words than in spoken nonwords. Percept. Psychophys .19(5), 394–398.

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

Samuel, A.G., 2001. Knowing a word affects the fundamental perception of the sounds within it. Psychol. Sci. 12(4), 348–351.

Simpson, S.A., Cooke, M., 2005. Consonant identification in N-talker babble is a non-monotonic function of N (L). J. Acoust. Soc. Am. 118, 2775–2778.

Taft, M., Hambly, G., 1986. Exploring the Cohort Model of spoken word recognition. Cognition 22, 259–28.

Van Engen, K.J., Bradlow, A.R., 2007. Sentence recognition in native- and foreign-language multi-talker background noise. J. Acoust. Soc. Am. 121(1), 519–526.