INTERSPEECH 2009 BRIGHTON

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

Véronique Boulenger[1], *Michel Hoen*[2], *François Pellegrino*[1], *Fanny Meunier*[1]

[1] Laboratoire Dynamique du Langage, CNRS UMR 5596, Lyon, France
[2] Stem Cell and Brain Research Institute, INSERM U846, Lyon, France

Veronique.Boulenger@ish-lyon.cnrs.fr

Abstract

This study investigates speech comprehension in competing multi-talker babble. We examined the effects of number of simultaneous talkers and of frequency of words in the babble on lexical decision to target words. Results revealed better performance at a low talker number (n = 2). Importantly, frequency of words in the babble significantly affected performance: high frequency word babble interfered more strongly with word recognition than low frequency babble. This informational masking was particularly salient for the 2-talker babble. These findings suggest that investigating speech-in-speech comprehension may provide crucial information on lexical competition processes that occur in real-time during word recognition.

Index Terms: 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. Yet 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 therefore constitutes a challenge that scientists have to face given its major social impact. In this study, we aimed at identifying the cognitive processes that come into play during speech-in-speech comprehension by examining lexical competitions during word recognition in cocktail party situations [1].

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 [2, 6]. One example illustrating the influence of lexical knowledge on speech recognition is the fact that speech sounds in words are recognized more quickly than speech sounds in non-words [3, 4]. These 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 [5, 6]. 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 do occur [7, 8]: *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. In the context of speech-in-speech comprehension, this 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 may particularly arise when only a few simultaneous talkers are present in the babble [9-11]. In a recent study [9], we examined the differential effects of acoustic-phonetic and lexical content of babble or babble-like noise (reversed speech) on word identification. Our results revealed that at a low talker number (N = 4), performance decreased in the natural babble compared to the reversed babble condition, suggesting increased informational masking and increased lexical competition due to the availability of identifiable lexical items from background.

The present study aimed at further breaking down informational masking into its different parts by examining real-time *lexical competitions* between target and background in speech-in-speech comprehension situations. We sought to assess whether and how frequency of words in the babble influences lexical access to target words. Word frequency has been widely shown to affect word recognition: high frequency words in a language are better recognized than low frequency words [12, 13]. Here, we hypothesized that if listeners are sensitive to linguistic (lexical) factors in the background babble, lexical competition between target and babble should vary depending on the frequency of words in this babble. In a lexical decision task, healthy participants were asked to decide whether target items mixed with multi-talker babble were words or not. The number of competing talkers (2, 4, 6 or 8) and the frequency of words in the babble (high or low) were manipulated. Our prediction was that informational masking should vary according to the number of talkers, with maximal informational lexical masking being reached at a low talker number (N = 2). We further hypothesized 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.

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* [14]. 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). Words in both lists were matched for length in letters and number of syllables. These lists were used as the babble signals. Eight French native speakers (50/50 female/male) recorded the two lists in a sound-attenuated room (sampling rate 44 kHz, 16 bit accuracy). Order of words in the lists was randomized and different for each talker. 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. Individual sources were then mixed into 2-, 4-, 6- or 8- mixed (half female half male) talker babble containing either high- or low-frequency words. Eight cocktail-party sound tracks were thus created (4 talker numbers x 2 frequencies of words in the babble): T2F+, T2F-, T4F+, T4F-, T6F+, T6F, T8F+ and T8F-.

2.2.2. Target words and pseudo-words

One hundred and twenty mono-syllabic, tri-phonemic French words were selected from *Lexique 2* [14] in a middle range of frequency of occurrence (mean = 53.89 p/m, SD = 69.77). One hundred and twenty mono-syllabic pseudo-words were also constructed by changing phonemes order from target words. All pseudo-words respected the phonotactic rules of French language. Stimuli were recorded in a sound-attenuated booth by a female French native speaker.

2.2.3. Stimuli and word lists

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 and 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. In the end, each list was made up of 120 stimuli, 15 appearing in each of the 8 babble conditions. Across lists, all target words were presented against the 8 multi-talker babble.

2.2.4. 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 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 advance from trial to trial. They could listen to each stimulus no more than once. Before the testing phase, they were given 16 practice items to accommodate themselves to stimulus presentation mode and target voice. To ensure that they paid attention to the task, participants could be asked to transcribe the word they had just heard on a piece of paper.

2.2.5. Statistical analysis

Reaction times (RT: time-interval between the onset of the target stimulus and the button press; in milliseconds) and correct identification rates (%) for target word identification were measured. Trials for which participants made no response or made mistakes (word response for a pseudo-word or vice-versa) were considered as errors and were not included in the RT analysis. Individual RT and identification rates were used as dependant variables in separate statistical analyses. A two-way repeated measures analysis of variance (ANOVA) was conducted, with talker number (2 vs. 4 vs. 6 vs. 8) and frequency of words in the babble (F+ vs. F-) as within-subjects (F1) and within-items factors (F2).

3. Results

3.1. Word RT analysis

The ANOVA revealed a significant main effect of talker number on participants' RT for target word identification ($F1 (3, 29) = 3.64$, $p = .015$; $F2 (3, 303) = 2.77$, $p = .041$). Mean RT were faster when 2 talkers were present in the babble (1000 ms ±91) compared to the 4- (1022 ms ±102), 6- (1020 ms ±87) and 8-talker conditions (1033 ms ±92; Figure 1). These three conditions did not significantly differ from each other. A significant main effect of frequency of words in the babble also emerged ($F1 (1, 31) = 6.51$, $p = .016$; $F2 (1, 101) = 6.22$, $p = .014$), with slower mean RT when words in the babble were highly frequent (1025 ms ±94) compared to when they were less frequent (1012 ms ±92; Figure 2). No interaction was observed between the two factors.

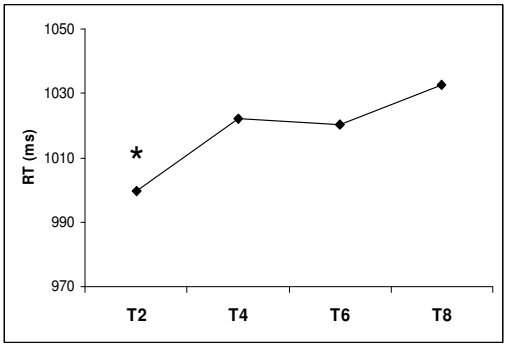

Figure 1: *Mean reaction times (ms) for target word identification when 2- (T2), 4- (T4), 6- (T6) and 8-talkers (T8) were present in the babble. * indicates a significant difference between T2 and other conditions.*

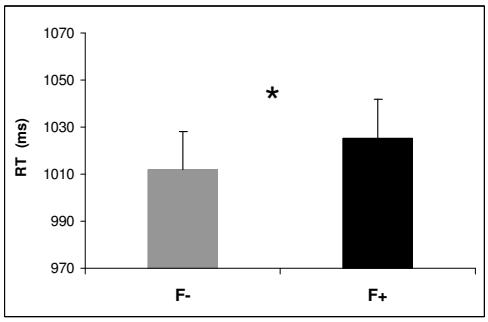

Figure 2: *Mean reaction times (ms) for target word identification when babble was composed of low frequency words (F-) and high frequency words (F+). Standard errors are reported.* indicates a significant difference between the conditions.*

Interestingly, RT distribution by items for target word identification was bimodal with about half words (53/120) being responded to in less than 1000 ms (mean = 941 ms ±38) while the other half (67/120) led to longer RT (> 1000 ms; mean = 1099 ms ±61). In order to examine whether type of background babble equally affected these two sets of items, we performed the two-way repeated measures ANOVA on these two sets separately. Whereas no main effects of talker number and of babble word frequency were found for target words with short RT, these two factors influenced performance for words with longer RT (> 1000 ms). First, a significant effect of talker number (F1 (3, 29) = 3.12, p = .029; F2 (3, 156) = 3.26, p = .029) emerged: RT were shorter in the 2-talker condition (1064 ms ±107) compared to the 4- (1101 ms ±140), 6- (1111 ms ±130) and 8-talker conditions (1112 ms ±132) which gave similar results. Second, frequency of words in the babble had a significant effect on performance (F1 (1, 31) = 5.64, p = .024; F2 (1, 52) = 9.86, p = .003), RT to target words being longer when babble was composed of high frequency words (F+; 1110 ms ±133) compared to low frequency words (F-; 1084 ms ±123). No interaction was found between the two factors.

To further investigate the effect of word frequency, we then compared performance within each babble condition with paired *t*-tests. Results revealed that mean RT were significantly longer in the F+ than in the F- condition when 2 talkers (F1 (1, 31) = 4.8, p = .036; F2 (1, 65) = 4.01, p = .049) and 8 talkers were present in the babble (F1 (1, 31) = 3.85, p = .058; F2 (1, 57) = 6.75, p = .012; Figure 3). No significant difference was observed for the two other babble conditions.

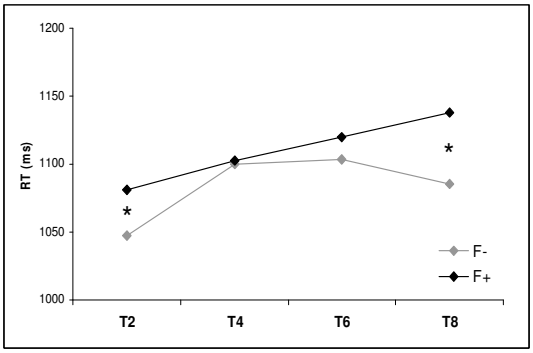

Figure 3: *Direct comparison of the effect of frequency of words in the babble (F- vs. F+) on mean reaction times (ms) to target words in each of the 4 multi-talker conditions. Only results for target items with longer RT (> 1000 ms) are reported. Note the significant difference (*) between F- and F+ in the 2- and 8-talker conditions.*

3.2. Word identification rate analysis

The ANOVA conducted on word identification rate revealed a significant main effect of talker number (F1 (3, 29) = 11.04, p < .001; F2 (3, 117) = 6.55, p < .001). Words were better recognized when 2 talkers were present in the babble (77 % ±13) compared to the 4- (70 % ±12), 6- (72 % ±14) and 8-talker conditions (70 % ±14) which did not significantly differ from each other. We also observed a significant main effect of frequency of words in the babble (F1 (1, 31) = 5.76, p = .023; F2 (1, 119) = 3.97, p = .049), with worse identification when words in the babble were highly frequent (71 % ±14) compared to when they were less frequent (74 % ±13). No interaction was observed between the two factors.

As for the RT analysis, we then performed the statistical analysis on the two sets of target words depending on mean RT (< 1000 ms or > 1000 ms). No significant effects of talker number and of babble word frequency were found for items with RT < 1000 ms. By contrast, we observed that the number of talkers significantly influenced performance for words with longer RT (F1 (3, 29) = 4.92, p = .003; F2 (3, 64) =, p = .004): identification rate was better in the 2-talker condition (73 % ±17) compared to the three other conditions (T4: 62 % ±18, T6: 65 % ±20 and T8: 63 % ±19). Contrary to what was found for RT, no significant effect of frequency of words in the babble was observed (F1 (3, 29) = 4.92, p = ns; F2 (1, 66) = 4.93, p = .076).

4. Discussion

This study investigated lexical competitions during speech-in-speech comprehension using a lexical decision task. Results revealed better recognition of target words both in terms of reaction times and identification rate when only 2 talkers were present in the babble compared to 4, 6 and 8-talker conditions. 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 [8-10]. 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 [8] 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 [9]. 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 [15, 16]. This phenomenon has been viewed as informational masking occurring at the acoustic–phonetic level [9].

Importantly, the study also demonstrated that 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. This was particularly observed for the 2- and 8-talker babble conditions and when only target words that led to longer RT (> 1000 ms) were considered. This latter finding may indicate that competitions between target and masker are more likely to affect speech recognition when target words are more difficult to recognize (i.e. lexical access is more difficult) than when response is more immediate. Assessing why this is so however remains an open question to address.

In the 2-talker 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 [17]. This linguistic interference may even be stronger when words in the babble are highly-frequent, as was observed here, since words in such babble may strongly activate the mental lexicon, thus 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 [2, 5, 6]. In the context of speech comprehension in high frequency word babble, this competition may be maximal, thus "overloading" the language processing system: lexical access to target words would be more difficult, resulting in worse performance for target word recognition (lengthened reaction times and increased error rates). It is also plausible that high-frequency words in the babble capture more the listener's attention than low frequency words, consequently reducing cognitive resources that are available to process target speech. With high frequency words in the babble, attention may indeed switch back and forth between target speech and babble so that word identification becomes more difficult.

The effect of frequency of words in the babble in the 8-talker condition is by contrast unlikely to result from 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) [9, 17]. In order to test this hypothesis, we performed an acoustic analysis following the method developed by Hoen et al. [9], which aimed at evaluating the effect of spectro-temporal saturation on speech comprehension. In agreement with our previous findings [9], results revealed that an increase in the number of simultaneous talkers monotonically increased saturation as measured by cepstral variation among segments ($p < .001$). Furthermore, this cepstral variation was significantly larger for the high frequency than for the low frequency word babble only in the 8-talker condition ($p = .0014$). In other words, when 8 talkers were present in the babble, acoustic/energetic features distinguished between high frequency and low frequency word babble. The effect of word frequency on target word recognition that we observed in this type of babble may accordingly stem from such acoustic differences rather than from lexical content of the babble.

Although future work needs to be done to further investigate the different types of linguistic interference – both at a low-level and a high-level of information – elicited by babble background, this study demonstrates that multi-talker babble can compete with speech recognition at the lexical level provided that the number of interfering talkers is rather low. Regarding psycholinguistic models of language processing, our results suggest that investigating the mechanisms underlying speech-in-speech comprehension may constitute an original approach to assess interactions between the different levels of speech processing and competition between these levels.

5. Conclusions

The present study revealed that speech comprehension in multi-talker babble triggers competitions at the lexical level between target and background: lexical factors from the babble such as word frequency can interfere with target word recognition. 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 background. Such findings highlight the importance of examining speech-in-speech comprehension situations as it could provide crucial information on competitive mechanisms that occur during language processing.

6. Acknowledgements

We would like to thank Emmanuel Ferragne for precious help in constructing the materials of the experiment. This project is carried out with financial support from the European Research Council (SpiN project to Fanny Meunier).

7. References

[1] Cherry, E., "Some experiments on the recognition of speech, with one and two ears", J Acoust Soc Am, 25: 975-979, 1953.

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

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

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

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

[6] McClelland, J.L. and Elman, J.L., "The TRACE model of speech perception", Cogn Psychol, 8, 1-86, 1986.

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

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

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

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

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

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

[13] Taft, M. and Hambly, G., "Exploring the Cohort Model of spoken word recognition", Cognition, 22: 259-282, 1986.

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

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

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

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