Effect of spatial separation on speech-in-noise comprehension in dyslexic adults

Marjorie Dole 1, Michel Hoen2, Fanny Meunier3

1 Laboratoire Dynamique du Langage, Lyon, France
2 Stem-Cells and Brain Research Institute, Bron, France
marjorie.dole@univ-lyon2.fr

Abstract
This study tested the use of binaural cues in adult dyslexic listeners during speech-in-noise comprehension. Participants listened to words presented in three different noise-types (Babble-, Fluctuating- and Stationary-noise) in three different listening configurations: dichotic, monaural and binaural. In controls, we obtained an important informational masking in the monaural configuration mostly attributable to linguistic interferences. This was not observed with binaural noise, suggesting that this interference was suppressed by spatial separation. Dyslexic listeners showed a monaural deficit in Babble, but no deficit of the binaural processing, suggesting compensation based on the use of spatial cues.

Index Terms: dyslexia, speech-in-noise, informational masking, binaural processing

1. Introduction
In everyday life, we are confronted with situations in which speech sounds must be segregated from concurrent noise. Speech-in-speech comprehension is a situation of first relevance for who is interested in speech perception.

Two types of masking occur in this situation: energetic masking (EM) which is produced when the target signal and interfering noise partly overlap in time and frequency at peripheral level; and informational masking (IM) which occurs when information inside the target signal and the interfering noise is of comparable nature. IM is thought to operate at a higher level. Typically, IM occurs when the competing flows are speech sounds. In such situations, IM is particularly important and is made up of different levels of linguistic information, creating competition between phonological, lexical and semantic cues [1], [2]. In speech-in-speech situation, both EM and IM occur, although EM represents only a small amount of the overall masking [3].

In a speech-in-speech situation, segregation between competing sounds can be facilitated by different cues, one of them being spatial separation: when target and competing noise originate from the same location, noise greatly affects the intelligibility of the target; however when the noise-source is moved away from the target its intelligibility is increased. This is reflected in the binaural release of masking, which we will define as the observed improvement of intelligibility of an ipsilateral target when the noise-source is progressively moved towards the contralateral ear [4], [5]. This effect is due to binaural cues (interaural level differences, ILD, and interaural time differences, ITD) which help differentiating the two competing flows.

Evaluating the use of these binaural cues can be useful to detect potential deficits in auditory processing in specific populations, one of these being dyslexic people. Developmental dyslexia is defined as a severe and long-lasting deficit in the acquisition of written language abilities, in the absence of neurological or sensory deficiencies. Causes of this impairment are still debated, one of the most popular hypothesis being a deficit in the representation and/or use of phonological information. This trouble could rely on a deficit of speech perception, which is corroborated by the fact that impairments have been shown in the perception of speech in concurrent noise [6], and in the representation of speech-in-noise sounds measured at the neural level [7], [8].

One possible explanation for this speech-in-noise deficit could be that dyslexics experience difficulties in the processing of binaural cues that are crucial to segregate between sound sources. McAnally & Stein (1996), for example, found reduced binaural level masking differences, a phenomenon that depends of the interaural phase relations of the signal and the concurrent noise [9]. Similarly, Edwards et al. (2004) used a fluctuating ITD to evaluate the use of ITD to extract a signal from background noise in dyslexic children [10]. They found a reduced ability to use ITD to localize the signal compared to controls. These results, with those obtained by Dougherty (1998), suggest a low-level deficit in the ability to use fine binaural temporal cues for the signal-to-noise extraction [11]. However, several other investigations failed to demonstrate such a deficit [12], [13].

It should be noted that all these studies investigate a deficit based on the detection of a tone in noise. As dyslexic listeners seem to be more affected by the background noise that is composed of speech [14], it is of great interest to evaluate how binaural advantage is affected by dyslexia in a speech-in-speech situation. Moreover, to our knowledge, no investigation has been conducted on the processing of ILD in dyslexic listeners.

We started addressing this issue with the investigation on the advantage due to spatial separation of speech and noise in normal readers [15]. In order to separate the different types of information necessary to create this unmasking effect, we used three types of noise: a babble noise, producing both EM and IM, a fluctuating speech-shaped noise, and a stationary broadband noise producing only EM. These noises were tested in three listening configurations: dichotic, monaural, and a binaural configuration with a monaurally presented target and a binaurally presented noise. When comparing the impact of babble compared to, respectively, speech-shaped and broadband noise on intelligibility scores in monaural configuration, we observed an important informational masking, suggesting that high level cues like phonetic and lexical information were predominant in the masking effect in comparison with low-level cues like spectral or temporal information. We also obtained a binaural unmasking effect, the amplitude of which being more important in babble-noise, suggesting that participants used the perception of location of target and noise in the case of informational masking, which was not the case in energetic masking.

The purpose of the present study is to pursue these investigations in dyslexic adults, with normal readers as controls.

We expect dyslexic participants to be more affected by the presence of background noise than normal readers, with particularly low performances in a speech masker. If we obtain
better performances in the configuration where noise is binaurally presented in control listeners than in dyslexic ones, we can conclude that this deficit partially relies on a low-level deficit of the processing of binaural cues. If the deficit is observed in the monaural configuration, this will suggest a more central deficit.

2. Method

2.1. Participants and procedures

Twenty volunteers with a history of dyslexia (mean age: 24.6, S.D. = 5.79) and 20 normal readers selected to match the dyslexics in age (mean age: 24.4, S.D. = 5.74) participated in this experiment. All were right-handed French speakers and had audiometric pure-tones thresholds ≤20dB HL at frequencies in the range of 250-8000Hz. They were asked to listen to auditory stimuli, delivered via headphones. Stimuli were dysyllabic words, presented in three types of noise. Subjects were asked to repeat the word they heard to the experimenter.

2.2. Stimuli

2.2.1. Target words

The target words were 126 dysyllabic words selected in a middle range of frequency of occurrence (ranging from 0.13 to 338.19, mean: 16.82, S.D.: 43.74) according to the French database Lexique2 [16]. They were pronounced by a 24-year-old French native female speaker and recorded in a sound-proof room.

2.2.2. Noises

Three types of noise were used: a babble-noise (Babble), a fluctuating speech-shaped noise (SSN) and a stationary broadband noise (BBN).

Babble was made up of 4 voices (2 male; 2 female); every single voice was recorded in a sound-proof room, reading extracts of French press. Individual recordings were modified according to the following protocol: 1) removal of silences and pauses of more than 1s; 2) suppression of sentences containing pronunciation errors, exaggerated prosody or proper nouns; 3) noise reduction optimized for speech signals, 4) intensity calibration in dB-A and normalization of each source at 80dB-A; 5) final mixing of individual sources into cocktail-party soundtracks.

The fluctuating speech-shaped noise (SSN) was made to have spectro-temporal characteristics equivalent to that of the Babble; to do so, we took our babble-noise and extracted envelope information below 60Hz. Using Fast Fourier Transformation (FFT), the power spectrum and phase distribution of the original signal were computed and the original phase information was discarded by randomizing phase distribution. An inverse FFT was used to generate a new signal with equivalent power spectrum but randomized phases convolved with the temporal envelope of the original babble-noise. Finally, the root mean square (rms) powers of the original and new signals were equalized.

Similarly, a stationary broadband noise (BBN) with spectral composition identical to our babble-noise was generated.

2.3. Stimuli and word lists

Stimuli consisted in a list of 126 dysyllabic words. They were composed by mixing each target word with randomly selected 4s chunks of noise; target words were inserted 2.5s after the noise onset.

Three presentation configurations were tested: i) dichotic, with target word in ipsilateral ear and noise in contralateral ear (Dichotic); ii) monaural, with target and noise in the same ear (Monaural) and iii) a binaural configuration, with target in ipsilateral ear and the same noise in the two ears (Binaural). This latter configuration created an effect of binaural release of masking in comparison with the monaural configuration.

For each configuration, the 3 noises were tested, leading to 9 experimental conditions, and 14 words were presented per condition.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Target ear</th>
<th>Contralateral ear</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dichotic</td>
<td>Speech</td>
<td>Noise</td>
<td>BBN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Babble</td>
</tr>
<tr>
<td>Monaural</td>
<td>Speech &amp; Noise</td>
<td>Silence</td>
<td>BBN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Babble</td>
</tr>
<tr>
<td>Binaural</td>
<td>Speech &amp; Noise</td>
<td>Noise</td>
<td>BBN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SSN</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Babble</td>
</tr>
</tbody>
</table>

Stimuli were presented at an intensity of 65 dB SPL and a SNR of 0 dB in the target ear and an intensity of 45 dB SPL in the contralateral ear (ILD = -20dB).

Nine lists of stimuli were created, so that across lists each word was presented in each condition and every noise. Within lists, frequency of words was counterbalanced.

Participants heard two blocks of 63 stimuli. In one block, the target word was presented in the right ear, and in the other block the target word was presented in the left ear. Order of presentation of blocks was randomized.

3. Results

Participants’ responses were analysed by calculating the proportion of words that corresponded to the target. These identification rates were converted to rationalized arcsine transform units [17]. RAU scores were calculated for each of the nine conditions of stimulation, and used as dependant variables in a 4-way repeated-measures analysis of variance (ANOVA), with population (Population) as inter-subject factors, and target ear (Ear), noise (Noise), and presentation configuration (Configuration) as intra-subject factors.

This analysis revealed a significant effect of Noise (F(2,76)=8.06, p<.001); planned comparisons revealed that performances in Babble differed from both performances in BBN (F(1,38)=9.93, p<.005) and SSN (F(1,38)=11.57, p<.005).

We found no main effect of Ear (F(1,38)<1, n.s.) but a main effect of Configuration (F(2,76)=263.21, p<.0001); planned comparisons revealed that all configurations differed from each other, with 97.1% of correct responses in dichotic configuration and 69.4% of correct responses in monaural configuration (F(1,38)=621.81, p<.0001). We observed a binaural unmasking effect, with better performances in binaural configuration (78.4% of correct responses) than in monaural configuration (F(1,38)=40.47, p<.0001) (Figure 1).

A significant interaction between these two latter factors (Noise and Configuration) was also noticeable (F(4,152)=6.53, p<.0001), as shown in Figure 2. Planned comparisons performed for the dichotic and binaural configurations showed...
that performances obtained in the three noises were not significantly different. However in monaural configuration performances for Babble were significantly worse than for BBN (F(1,38)=14.77, p<.001) and SSN (F(1,38)=31, p<.0001).

![Figure 1: Identification rates as function of the type of Configuration](image1)

In order to test the effect of binaural unmasking as a function of the type of Noise, we also performed planned comparisons between binaural and monaural configurations for each noise. Results indicated a significant binaural unmasking effect for BBN (F(1,38)=9.22, p<.005) and Babble (F(1,38)=47.75, p<.0001) but not for SSN (F(1,38)=2.29, p=.14).

![Figure 2: Identification rates as a function of Configuration and Noise](image2)

The 4-way ANOVA also revealed a main effect of Population (F(1,38)=5.94, p<.05), showing that global performances were better for Controls (83.4%) than for dyslexics (79.8%).

The Population*Noise interaction failed to reach statistical significance (F(2,76)=2.3, p=107), but as we were specifically interested in comparing populations on babble-noise, we performed planned comparisons between the two populations for each noise. We obtained no effect of Population for BBN (F(1,38)<1, n.s.) but dyslexic listeners were significantly more disturbed by SSN (F(1,38)=6.12, p<.05) and Babble (F(1,38)=5.67, p<.05) than controls.

We also observed a significant interaction Population*Configuration*Noise (F(4,152)=3.52, p<.01) (Figure 3). Planned comparisons revealed that in BBN, there was no significant difference between the two populations regardless of Configuration (Dichotic: F(1,38)<1, n.s.; Monaural: F(1,38)<1, n.s.; Binaural: F(1,38)<1, n.s.). For SSN, populations did not differ for Dichotic configuration (F(1,38)<1, n.s.) neither Monaural (F(1,38)<1, n.s.) but for Binaural configuration Dyslexics had lower performances than Controls (F(1,38)=8.65, p<.01). We therefore performed planned comparisons to determine the binaural unmasking effect in the two populations separately, and observed that dyslexic participants did not show any binaural unmasking for SSN (F(1,38)<1, n.s.), whereas Controls showed this effect (F(1,38)=5.05, p<.05).

In babble-noise, Controls had better performances than Dyslexics in dichotic (F(1,38)=8.13, p<.01) and monaural configurations (F(1,38)=15.03, p<.001), but no difference was observable for the binaural configuration (F(1,38)<.0001, n.s.)

Finally amplitude of unmasking was calculated by subtracting, for each noise, performances obtained in the monaural configuration from those obtained in the binaural configuration. A two way ANOVA with Population as inter-subject factor and Noise as intra-subject factor was performed on this difference and revealed no main effect of Population (F<1, n.s.) but a main effect of Noise (F(2,76)=8.78, p<.0005); planned comparisons revealed that Babble produced significantly more release of masking than BBN (F(1,38)=8, p<.01) and SSN (F(1,38)=15.31, p<.0005). A significant Population*Noise interaction (F(2,76)=3.37, p<.01) was also observed. Further analyses revealed that populations did not differ for BBN (F(1,38)=1, n.s.), nor for SSN (F(1,38)=2.77, n.s.), but the amplitude of the unmasking effect in Babble was significantly more important for the Dyslexics than for Controls (F(1,38)=7.72, p<.01).

![Figure 3: Identification rates as a function of noise and configuration in the two populations](image3)

4. Discussion

In this study we further investigated the deficit present in dyslexic adults in speech-in-noise comprehension, by evaluating binaural release of masking in dyslexic and normal listeners in energetic and informational masking.

In Controls, we obtained in the dichotic configuration nearly 97% of correct responses; indeed, the spatial separation between target and competing noise is maximal, so participants had no difficulty in detecting the target. In the monaural configuration however, we observed an important decrease of performances, because target and background are not spatially separated so the segregation between the two sources became much more difficult. As previously shown [15], this masking was more important in Babble, which indicates that linguistic cues are more important than spectral or temporal ones to create this masking effect.

In the binaural configuration, we obtained a release of masking. In this condition, subjects used the interaural cues available (ILD) to allow segregation between noise and target. These ILD allowed participants to create an image of the noise
inside their heads which was a little more central than the target, and so improved intelligibility. An interesting finding is that this unmasking effect was dependent of the type of background noise. Indeed, the unmasking amplitude was more important for informational than for energetic masking. This is due to the fact that spatial unmasking for energetic and informational masking relies on two different mechanisms [18]: in the case of energetic masking, target and speech are different and easy to disentangle. The use of spatial localization as a cue was not necessary and unmasking only resulted from the processing of the ILD. In the case of informational masking, however, target and interferer are much more difficult to separate, because of their shared linguistic nature; thus participants need to use perceived spatial location to separate these two flows. Thus unmasking resulted much more from perceptual factors.

In this study, in agreement with the literature, we observed a significant deficit in dyslexic participants for the speech-in-noise perception [6]. Moreover, our results showed a trend towards a worsening of this deficit when background is composed of speech. The question is whether this deficit relies on a low-level deficit of the processing of the ILD or on a more central deficit.

For the Fluctuating noise, we observed a deficit in the binaural configuration for dyslexics in comparison with normal readers. This noise contained no linguistic information and was only composed of spectro-temporal envelope of the Babble noise, so it produced only energetic masking, and separation in this case was only based on the processing of the ILD. This result suggests a low-level deficit in the processing of the ILD for a target. However, this processing was not altered in stationary noise, which shows that dyslexic listeners experienced some difficulties in the processing of the temporal information when this one is binaurally presented.

For babble-noise, our results clearly showed a deficit in the monaural configuration; dyslexic participants were more disturbed by the background when this one was speech. This monaural effect was not present for the two other noise types, we can so conclude that it is the linguistic cues present in the babble-noise that are particularly disturbing. In this frame of mind, we can note that dyslexic participants seem to be also more disturbed than normal readers by babble-noise when this one is presented in a dichotic configuration. In the binaural configuration however, no differences between Dyslexics and Controls emerged. This suggests that dyslexics experienced no deficit for localization of the two auditory flows in order to separate them, and on contrary compensate the deficit created by the speech masker by using the perceived location. Together, these results suggest a more central deficit for the processing of linguistic competitions rather than a deficit of spatial localization.

5. Conclusion
This study was focused on the speech-in-noise deficit observed in dyslexic listeners. As deficits in the processing of binaural cues have been observed, we decided to focus our attention on the processing of ILD in different types of masking. Our results showed impaired processing of the linguistic competitions with a decrease of performances in a monaural speech-in-speech situation, but no deficit of the use of perceived localization to separate auditory sources. When background is composed of a fluctuating noise, no unmasking was observable, suggesting an absence of the processing of the ILD. These results suggest both a low level deficit of the processing of the ILD in the case of energetic masking, and a more central deficit in the case of informational masking.

6. Acknowledgements
The first author’s thesis research is funded by a PhD grant from the “Cluster régional de recherche: Handicap Vieillissement et Neurosciences (HVN)” from the Rhône-Alpes region, France. We are grateful to the European Research Council for funding this project via an ERC starting grant attributed to F. Meunier (SpiN project).

7. References