JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

French version of the coordinate response measure corpus and its validation on a speech-on-speech task

Vincent Isnard,[a] **Véronique Chastres, and Guillaume Andéol**
Institut de Recherche Biomédicale des Armées, Brétigny-sur-Orge, France
vincent.isnard@def.gouv.fr, veronique.chastres@def.gouv.fr, guillaume.andeol@def.gouv.fr

Abstract: Since its creation, the coordinate response measure (CRM) corpus has been applied in hundreds of studies to explore the mechanisms of informational masking in multi-talker situations, but also in speech-in-noise or auditory attentional tasks. Here, we present its French version, with equivalent content to the original version in English. Furthermore, an evaluation of speech-on-speech intelligibility in French shows informational masking with similar result patterns to the original data in English. This validation of the French CRM corpus allows to propose the use of the CRM for intelligibility tests in French, and for comparisons with a foreign language under masking conditions.

[Editor: Matias Zanartu] https://doi.org/10.1121/10.0028059

Received: 1 March 2024 **Accepted:** 8 July 2024 **Published Online:** 25 July 2024

1. Introduction

The coordinate response measure (CRM) corpus is a corpus of recorded sentences dedicated to speech intelligibility tests, initially intended for a military and operational context. Indeed, the creation of this corpus was carried out mainly at the U.S. Air Force Research Laboratory (Ohio), and aimed to test communications performance in noisy environments and with communication devices altering the quality of the audio signal (Moore, 1981). In order to systematize these evaluations, Bolia *et al.* (2000) therefore recorded English sentences with all the same structure and pronounced by male and female talkers. This corpus was tested and validated by Brungart (2001) for speech-on-speech and speech-in-noise masking conditions.

The advantages of this corpus are multiple and its scope goes beyond military research. Indeed, it presents 256 controlled and standardized utterances recorded by eight male and female talkers, the recordings are proposed in a standard format, it is easily understandable by the participants and the structure of the sentences allows to adjust the participants' task according to different experimental contexts. For all these reasons, this corpus has been used and analyzed in a very large number of studies: in 2023, Google Scholar indicates 624 citations to the article by Bolia *et al.* (2000) and 1027 citations to the article by Brungart (2001). The CRM corpus is particularly appropriate for the study of informational masking in multi-talker situations, to highlight cognitive resources explaining intelligibility performances for "masking that occurs beyond energetic masking" (Kidd *et al.*, 2008). Yet, the use of the CRM corpus extends in particular to hearing status assessment [e.g., DiNino *et al.* (2022)], evaluation of listening effort with spatialized talkers [e.g., Lanzilotti *et al.* (2022)], or auditory attentional tasks [e.g., Gygi and Shafiro (2012)].

In order to extend the dissemination of the CRM corpus, authors have proposed versions in other languages: e.g., in British English (Kitterick *et al.*, 2010), in German (Schepers *et al.*, 2017), in Mandarin (Wang *et al.*, 2019), in Persian (Amiri *et al.*, 2020), in Dutch [a child-appropriate version of the CRM (Nagels *et al.*, 2021)], or in Spanish (Lelo de Larrea-Mancera *et al.*, 2023). However, no French version had yet been published. Yet, phonological and phonetic differences between the two languages could lead the talker segregation easier in English than in French. For instance, in English, there are more phonemes, the tonic accent is stronger and more variable between English talkers, and their rhythm of pronunciation could also be more variable (Markey, 1998; Selkirk, 2015).

Extending the availability of the CRM corpus in other languages allows the comparison of performances according to different native languages, but also to foreign languages, in particular to evaluate the cognitive resources involved in each situation [e.g., listening effort (Lanzilotti *et al.*, 2022)]. Here, we propose a version of the CRM corpus in French with content equivalent to that of the original version by Bolia *et al.* (2000). Then, we present speech-on-speech intelligibility results very similar to the original results in English, although obtained following a shortened version of the original protocol (Brungart, 2001).

[a]Author to whom correspondence should be addressed.

2. Corpus in French

2.1 Translation into French

The original corpus contains 2048 sentences all having the same structure: "Ready <call sign>, go to <color> <number> now" (e.g., "Ready baron, go to blue five now"). The participants' task is to indicate the color and number corresponding to a given call sign. The sentences are pronounced by eight talkers (four men, four women), with eight different call signs, four colors and eight numbers. In French, the wording of the sentence is as follows: "Prêt <call sign>, va au point <color> <number> go." Thus, the structure of the sentence is close to the original formulation, with an identical number of syllables.

The choice of call signs, colors and numbers in French is specified in Table 1. For the French version, a selection of call signs from the NATO phonetic alphabet commonly used in French was chosen. Only one call sign is identical between the two lists (Charlie). In French as in English, all call signs are made up of two syllables. The colors chosen in French are the same as in English except for the color "Jaune"/"Yellow," instead of "Blanc"/"White," in order to avoid phonetic redundancy with "Bleu"/"Blue." All colors are composed of one syllable. The numbers in French are the same as in English (between one and eight). In both versions, all the numbers are composed of one syllable except "Seven" in English which has 2.

2.2 Recording of the corpus

The talkers were 4 men and 4 women, aged between 22 and 25 years old ($M = 23.3 \pm 1.0$). They all were French-native talkers. No talker presented any particularity of accentuation or pronunciation and Standard French has always remained their main language in their daily life. All talkers signed a consent form to authorize the recording of their voices and the use of the recordings for research purposes.

The recording apparatus was similar to that described by Bolia *et al.* (2000). The talkers were placed in a double-walled IAC audiometric booth measuring $1.93 \times 1.83 \times 2.00$ m^3. Talker recording was carried out with a Brüel & Kjaer type 4165 1/2-in. microphone connected to a Brüel & Kjaer type 2669C preamplifier and a Brüel & Kjaer type 5935 power supply. The talkers were placed in front of the microphone less than 10 cm away. Outside the cabin, the analog-digital/digital-analog conversion of the sound signal was carried out using an RME Fireface UC sound card connected to a computer. A custom program on Max 8 (Cycling '74) provided the interface for recording, with a resolution of 16 bits and a sampling frequency of 48 kHz.

Before each sentence recording, the talkers listened to a standard sentence, previously recorded by one of the authors in order to regulate the speaking rhythm between the talkers. They had to press a key on a computer keyboard to start listening to this sentence, which was broadcast through a Fostex 6301B speaker, and displayed simultaneously on a computer screen in front of them. Then, they had to press the same key to start recording for a maximum duration of 3 s. Outside the cabin, the experimenter simultaneously listened to the recording through Beyerdynamic DT-770 headphones. After each sentence, the experimenter validated the recording or rejected it if the sentence was incomplete or pronounced at an inappropriate pace, in which case the talker had to restart the recording. The sentences were recorded one after the other entirely, successively going through the levels of the variables presented in Table 1 (call signs, then colors, then numbers). In addition, the recording of the first 32 sentences served as training, to gradually adapt the talker's speech in rhythm and prosody. They were re-recorded immediately following the recording of the 256 sentences for their final version. The total duration of the session was approximately 45 min per talker.

Once the corpus was recorded, it was edited and normalized similarly to Bolia *et al.* (2000), with a custom program on MATLAB R2022a (Mathworks). First, all audio files were filtered from 80 Hz to 8 kHz using an IIR bandpass filter of order 20. Then, all the beginnings of sentences were synchronized by automatically removing the silence at the beginning and at the end of the sentence. Then, all the waveforms were checked one by one, visually and aurally. For 22

Table 1. The CRM variables are indicated in their original version in English and in the version proposed here in French.

Call signs		Colors		Numbers	
English	French	English	French	English	French
Arrow	Alpha	Blue	Bleu	One	Un
Baron	Delta	Green	Vert	Two	Deux
Charlie	Charlie	Red	Rouge	Three	Trois
Eagle	Echo	White	Jaune	Four	Quatre
Hopper	Kilo			Five	Cinq
Laker	Oscar			Six	Six
Ringo	Tango			Seven	Sept
Tiger	Whisky			Eight	Huit

sentences out of all 2048 sentences recorded, silence was added at the beginning of the sentence in order to adapt the synchronization with the rest of the corpus. For 72 sentences, processing was carried out using AUDACITY software on short and isolated snippets, ensuring a natural rendering. Hissing noises were treated with the "Noise Reduction" function, and low-cut filtering with the "Graphic Equalizer" function, with a linear slope between 315 Hz (0 dB attenuation) and 125 Hz (20 dB attenuation). Finally, all sounds were normalized in amplitude to the same root mean square level. The average duration of the recordings is 2.2 ± 0.1 s.

3. Validation of the corpus in French

3.1 Methods

Ten new participants aged between 23 and 57 years old took part in the experiment (6 men and 4 women; mean age 34.0 ± 9.9). All participants were French-native talkers and gave written informed consent to participate in this experiment. They all self-reported normal hearing. They were placed in the audiometric booth and wore Beyerdynamic DT-770 headphones. The test was programmed on MATLAB with a graphical user interface presented on a computer screen, and which consisted of a 4 × 8 matrix of colored buttons. The buttons in each of the four rows were respectively colored in blue, yellow, red and green. In columns 1 to 8, a number was indicated on each button from 1 to 8.

The stimuli were composed of two mixed sentences from the CRM corpus in French, with random talkers. The call sign "Delta" was assigned to the target sentence. The masker sentence could take any other call sign randomly among the remaining seven call signs. The colors and numbers for both sentences were random, with different colors and numbers for the target and the masker. The target and the masking sentences were presented simultaneously under diotic listening. The sound level of the masking sentence was fixed, while the sound level of the target sentence was adjusted according to one of the 8 target-to-masker ratios (TMRs) randomly, between −9 and +12 dB in steps of 3 dB. In addition, the sound level of the mixture was randomly roved over a 6-dB range [in 1-dB steps (Brungart, 2001)]. For a TMR of 0 dB, the sentences were played at approximately 55 dBA, as measured with a Brüel & Kjaer type 2250 sound level meter and a Brüel & Kjaer type 4153 artificial ear connected to a G.R.A.S. Type 12AK power supply.

Participants completed one test block of 240 trials, thus on average around 30 repetitions per TMR. On each trial, participants had to indicate the color and number corresponding to the target call sign by clicking on the corresponding button. There was no time limit for responding, and the next trial started 0.5 s after their response.

3.2 Statistical analyses

The TMR factor had eight levels: from −9 to 12 dB TMR in steps of 3 dB; while the three levels of the masking condition factor were: "different sex," "same sex," and "same talker." Due to random sampling, for two participants, no trials were drawn at 12 dB TMR in the condition "same talker" (i.e., with the same talker in both target and masker sentences). These two missing values were replaced by the median of the values obtained from the other eight participants. We carried out several generalized linear models with TMR and masking condition as within-subject factors. Tukey-HSD *post hoc* tests were conducted when main effects or interactions were significant. Before running statistical tests, all required assumptions were checked, concerning normality of the data, variance homogeneity between groups and sphericity for within-subjects testing. Statistical analyses were performed using STATISTICA software version 13 (TIBCO Software, Inc., CA).

4. Results

4.1 Informational masking

Figure 1 shows the intelligibility performance for color and number, in the three masking cases: different sex, same sex, same talker. In particular, the analyses carried out for the correct recognition of both the number and the color show a significant main effect of the masking condition [$F(2, 18) = 36.153$, $p < 0.0001$, $\eta_p^2 = 0.801$], of the TMR [$F(7, 63) = 26.793$, $p < 0.0001$, $\eta_p^2 = 0.749$], as well as a significant interaction between both factors [$F(14, 126) = 6.804$, $p < 0.0001$, $\eta_p^2 = 0.431$]. As expected, intelligibility was better for "different sex" vs "same sex" [$p < 0.001$], and for "same sex" vs "same talker" [$p < 0.05$]. On average masking conditions, intelligibility was equivalent from −9 to 0 dB TMR [$p > 0.05$] and from 3 to 12 dB TMR [$p > 0.05$], except between 3 and 12 dB TMR [$p < 0.05$]. Yet, it was different between these two TMR areas, with overall better intelligibility scores for higher TMRs [$p < 0.01$]. Interestingly, the significant interaction between the TMR and the masking factors indicated that when the difficulty of the masking condition increased, intelligibility performances dropped around 0 dB TMR. Indeed, for the "different sex" condition, intelligibility decreased very slightly and monotonously with TMR, with a significant difference only between −9 and {3 to 12} dB TMR [$p < 0.01$; $p > 0.05$ for all other comparisons]. While for the "same talker" condition, the intelligibility score curve showed a local minimum at 0 dB TMR, significantly lower on one side from −9 dB TMR, and on the other from {3 to 12} dB TMR [$p < 0.001$].

As for the English corpus (Brungart, 2001), the distribution of responses confirmed the predominance of informational masking in this speech-on-speech task (Fig. 2). Indeed, the incorrect responses were mainly taken from the words of the masking sentence and not guessing responses.

Finally, note that the participants seemed to be subject to a moderate learning effect. For this analysis, the 240 trials performed by each participant were divided into 5 blocks of 48 successive trials. The effect of the learning factor

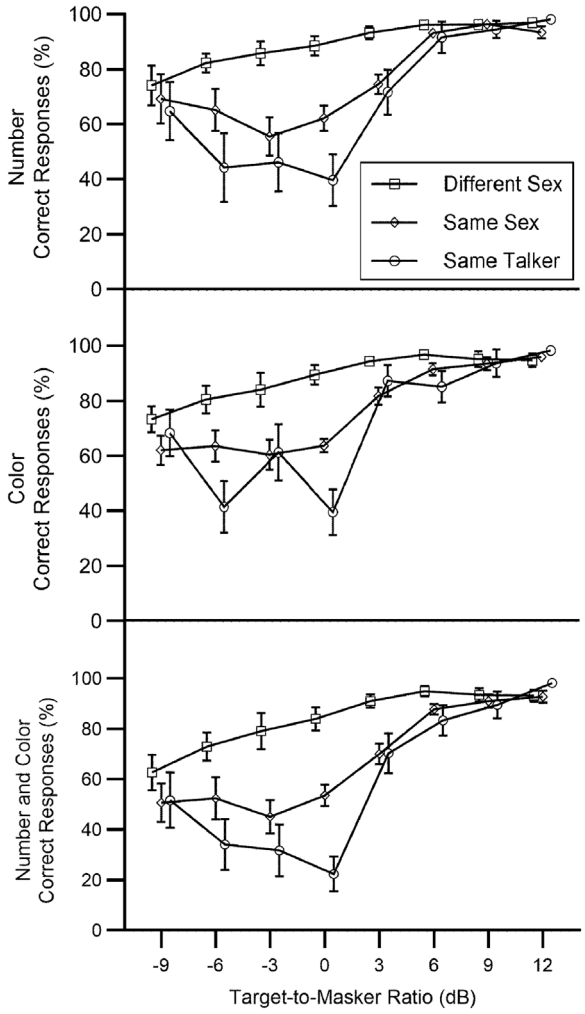

Fig. 1. Percentage of correct identifications as a function of TMR. The top panel shows the average score for number identification. The middle panel shows the average score for color identification. The bottom panel shows the average score for both the number and color identification. The data are shown separately for trials where the competing talkers were of different sexes, where the talkers were either male or female, and where the same talker was used for both the target and masker sentences. The error bars represent the standard error of the mean.

only indicated a non-significant trend [$F(4, 36) = 2.320$, $p = 0.076$, $\eta_p^2 = 0.205$], with an increase in average score after the first block which seemed to stabilize after the second or third block [respectively, (1) 70.21 ± 4.27%, (2) 75.42 ± 3.15%, (3) 76.46 ± 3.03%, (4) 79.79 ± 4.23%, (5) 78.13 ± 3.56%].

4.2 Intelligibility according to target and masking talkers

We found significant differences in the average score depending on the target and masking talkers [respectively, $F(7, 63) = 2.344$, $p < 0.05$, $\eta_p^2 = 0.207$; $F(7, 68) = 3.458$, $p < 0.01$, $\eta_p^2 = 0.278$] (Fig. 3). In a target sentence, the average scores for the talkers no. 1 and 6 were significantly lower than for the talkers Nos. 0, 2, and 7 [$p < 0.05$]. In a masking sentence, the average scores for the talker No. 4 were significantly higher than for the talkers Nos. 0, 3, 5, and 7 [$p < 0.05$], whereas the average scores for the talker No. 5 were significantly lower than for the talkers Nos. 1, 2, 4, and 6 [$p < 0.05$]. The average scores for the talker No. 2 were also significantly higher than for the talker No. 7 [$p < 0.05$]. However, there was no significant difference between sexes, in target or in masking sentences [respectively, $F(1, 9) = 0.089$, $p = 0.772$, $\eta_p^2 = 0.0098$; $F(1, 9) = 1.307$, $p = 0.283$, $\eta_p^2 = 0.1268$].

4.3 Intelligibility of colors and numbers

The average score varied depending on colors and numbers, and on whether they were presented in target or masking sentences (Fig. 4). First, as targets, the main effect on colors was significant, but not on numbers [respectively, $F(3, 27) = 8.121$, $p < 0.001$, $\eta_p^2 = 0.474$; $F(7, 63) = 1.359$, $p = 0.238$, $\eta_p^2 = 0.131$]. Indeed, the average score was significantly higher for the color "Vert" than for the other three colors [$p < 0.001$]. Second, as maskers, the average score was not

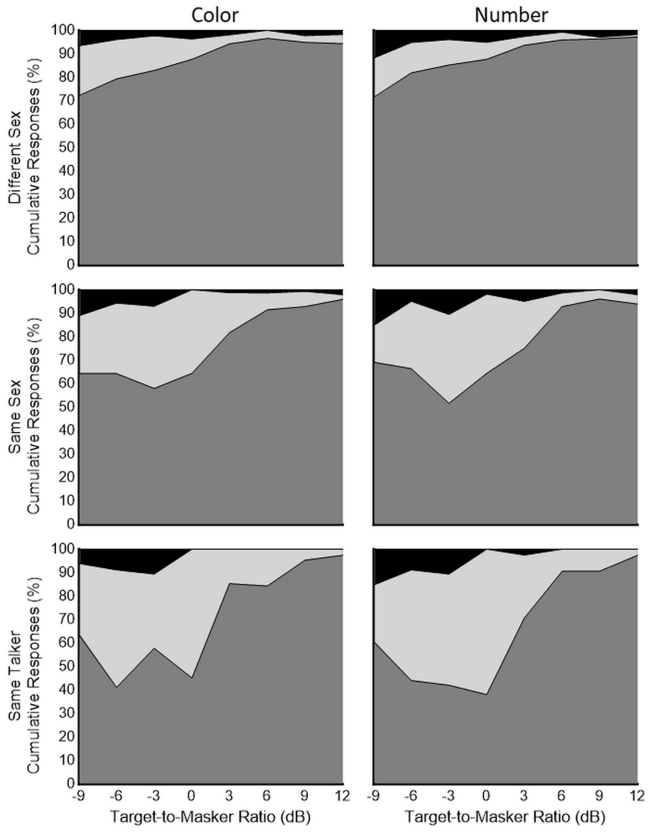

Fig. 2. Distribution of the listeners' responses among the correct word spoken by the target talker (dark gray), the incorrect word spoken by the masking talker (light gray), and all the other possible incorrect responses (black). The left and right panels show, respectively, the score for color and for number identification.

significantly different for colors, whereas it was different for numbers [respectively, $F(3, 27) = 1.476$, $p = 0.243$, $\eta_p^2 = 0.141$; $F(7, 63) = 3.012$, $p < 0.01$, $\eta_p^2 = 0.251$]. Indeed, the average score for the number 5 was significantly lower than for all other numbers except 7 [$p < 0.05$]. The average score for 7 was also lower than for 2 and 3 [$p < 0.05$].

5. Discussion

The French version of the CRM corpus is as close as possible to the original English version, e.g., structure of the sentences, same number of recorded sentences, same number of male and female talkers. For comparison, some versions of the CRM corpus in other languages are also very close to the original English version [e.g., Spanish version (Lelo de Larrea-Mancera *et al.*, 2023)], while other versions have been adapted to specific experimental contexts [e.g., only 2 talkers recorded instead of 8 in the German version (Schepers *et al.*, 2017)]. Although there are differences between the terms due to translation, for the call signs and for one color, their proximity of meaning and formulation allows a comparison of results for a given task.

In the speech-on-speech intelligibility task, we note that differences in performance exist in both French and English corpuses, respectively, depending on the talkers and the words spoken (e.g., in a target sentence, a better recognition of the color "Vert" in the French corpus, and "White" in the English corpus). The differences in average scores across the individual talkers are of the same order of magnitude as for the original English corpus, i.e., around 10%, without, however, one talker standing out from the others both as target and masker, unlike in Brungart (2001). Therefore, the variability of voice characteristics used to perform the task seems to be equivalent between the two versions of the corpus.

Some differences are also visible in the shape of the identification function in the "same talker" condition: in the original study a plateau was observed from adverse to 0 dB TMR whereas in the present study, a U shape was observed in the same range of TMR. In the "same talker" condition, the identification of the target talker is based on level cues. Therefore, adverse TMR with larger level differences between talkers could paradoxically facilitate the talker segregation, explaining the U shape. However, previous studies have shown that there are large individual differences in the listener's ability to use the level cues (Thompson *et al.*, 2015; Andéol *et al.*, 2017; Lanzilotti *et al.*, 2022). Therefore, it seems that in the original study the plateau could be explained by the mix of listeners able/not able to use level cues whereas in the present study most of the listeners were able to use level cues.

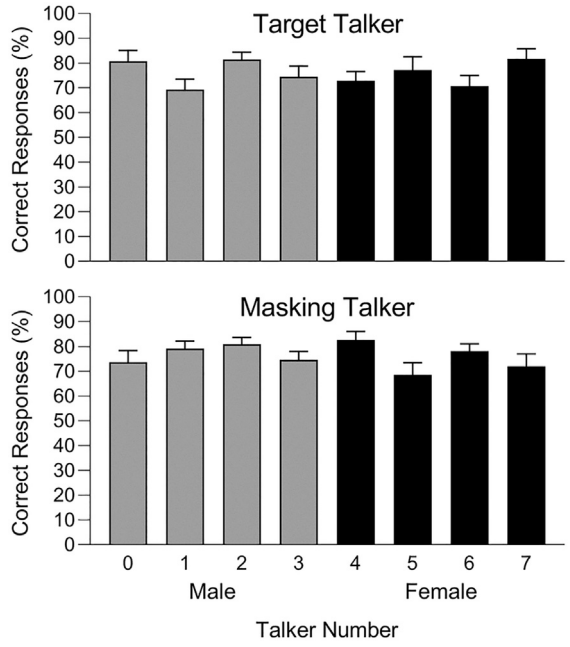

Fig. 3. Percentage of correct overall responses as a function of the talker used in the target sentence (top panel) and the masker sentence (bottom panel). The error bars represent the standard error of the mean.

Nevertheless, a clear pattern emerges from the overall data, with a distribution of listeners' responses obtained with the French corpus close to that observed with the English corpus in the original study (Brungart, 2001). In particular, it shows that informational masking occurs when the masker level increases: listeners answered the words pronounced by the masker. In case of energetic masking, guessing responses would be observed (Kidd *et al.*, 2008). Moreover, these intelligibility performance data are consistent with those of the English corpus, but for a reduced number of trials compared to the original English data. If we observed average scores that are sometimes lower for the French corpus compared to the

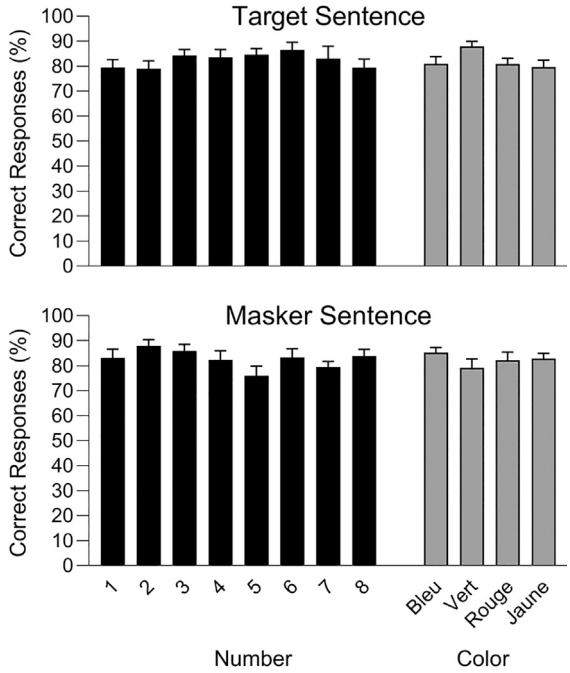

Fig. 4. Percentage of correct identifications as a function of the number or color word used in the target sentence (top panel) and the masker sentence (bottom panel). The data on the left and right show, respectively, the average score for number identification and color identification. The error bars represent the standard error of the mean.

US English one, in particular for the number recognition, these differences can also be explained by the proportion of trials linked to a learning effect of the task. Finally, the limited time per participant (around 20 min) allows us to consider uses of the CRM corpus in other experimental or clinical contexts with temporal constraints.

6. Conclusion

The French version of the CRM corpus has the same recording format as the original English version and therefore opens the possibility of using the CRM corpus to French-speaking participants. In the speech-on-speech intelligibility task, the patterns of results are similar to those obtained by Brungart (2001), with a predominant effect of informational masking when the level of masking increases. Thanks to the dissemination of versions of the CRM in other languages, this also opens the possibility of studying the interactions between intelligibility performance with other cognitive factors such as mastery of a foreign language or cognitive attentional abilities.

Acknowledgments

This work was supported by the French State, under the France 2030 Investment Plan financed by Bpifrance.

Author Declarations

Conflict of Interest

The authors state that they have no conflicts of interest to disclose.

Ethics Approval

This study has received approval from the French Ethics Committee (CPP Tours–Région Centre–Ouest 1, No. 2020T2-15). All participants provided written informed consent prior to any data collection.

Data Availability

The CRM corpus in French and the data collected for this study are available online at https://doi.org/10.5281/zenodo.11263054 (Last viewed 23 May 2024).

References

Amiri, M., Jarollahi, F., Jalaie, S., and Sameni, S. J. (**2020**). "A new speech-in-noise test for measuring informational masking in speech perception among elderly listeners," Cureus **12**, e7356.

Andéol, G., Suied, C., Scannella, S., and Dehais, F. (**2017**). "The spatial release of cognitive load in cocktail party is determined by the relative levels of the talkers," J. Assoc. Res. Otolaryngol. **18**, 457–464.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

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

DiNino, M., Holt, L. L., and Shinn-Cunningham, B. G. (**2022**). "Cutting through the noise: Noise-induced cochlear synaptopathy and individual differences in speech understanding among listeners with normal audiograms," Ear Hear. **43**, 9–22.

Gygi, B., and Shafiro, V. (**2012**). "Spatial and temporal factors in a multitalker dual listening task," Acta Acust. united Ac. **98**, 142–157.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (**2008**). "Informational masking," in *Auditory Perception Sound Sources*, Springer Handbook of Auditory Research Vol. 29 (Springer, Berlin), pp. 143–189.

Kitterick, P. T., Bailey, P. J., and Summerfield, A. Q. (**2010**). "Benefits of knowing who, where, and when in multi-talker listening," J. Acoust. Soc. Am. **127**, 2498–2508.

Lanzilotti, C., Andéol, G., Micheyl, C., and Scannella, S. (**2022**). "Cocktail party training induces increased speech intelligibility and decreased cortical activity in bilateral inferior frontal gyri. A functional near-infrared study," PLoS One **17**, e0277801.

Lelo de Larrea-Mancera, E. S., Solís-Vivanco, R., Sánchez-Jimenez, Y., Coco, L., Gallun, F. J., and Seitz, A. R. (**2023**). "Development and validation of a Spanish-language spatial release from masking task in a Mexican population," J. Acoust. Soc. Am. **153**, 316–327.

Markey, A. (**1998**). "A contrastive analysis of French and American English," M.A. dissertation, Department of Linguistics, Bryn Mawr College, Bryn Mawr, PA.

Moore, T. (**1981**). "Voice communication jamming research," in *AGARD Conference Proceedings: Aural Communication in Aviation*, Neuilly-Sur-Seine, France, pp. 2:1–2:6.

Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., and Başkent, D. (**2021**). "School-age children benefit from voice gender cue differences for the perception of speech in competing speech," J. Acoust. Soc. Am. **149**, 3328–3344.

Schepers, I. M., Beck, A. K., Bräuer, S., Schwabe, K., Abdallat, M., Sandmann, P., Dengler, R., Rieger, J. W., and Krauss, J. K. (**2017**). "Human centromedian-parafascicular complex signals sensory cues for goal-oriented behavior selection," Neuroimage **152**, 390–399.

Selkirk, E. O. (**2015**). *The Phrase Phonology of English and French* (Routledge, New York).

Thompson, E. R., Iyer, N., Simpson, B. D., Wakefield, G. H., Kieras, D. E., and Brungart, D. S. (**2015**). "Enhancing listener strategies using a payoff matrix in speech-on-speech masking experiments," J. Acoust. Soc. Am. **138**, 1297–1304.

Wang, Y., Lu, Z., Yang, X., and Liu, C. (**2019**). "Measuring Mandarin speech recognition thresholds using the method of adaptive tracking," J. Speech. Lang. Hear. Res. **62**, 2009–2017.