AUTHOR QUERY FORM

Author Please Answer All Queries

AQ1. For indexing purposes, please confirm that author names have been correctly identified as given names (blue) and surnames (red). Color in the byline will not appear on the final published version.

AQ2. Please provide the department/division name (if any) for affiliations and corresponding address.

AQ3. The term 'compared to' has been changed to 'compared with.' Please check and confirm if this is okay.

AQ4. Please provide a general caption for Figure 1 (for placement before the description of part labels).

AQ5. Tables or figures that are not the authors' own original work or are adapted from published content require permission letters in order to publish. If applicable, please supply the needed letters granting permission to publish these, or confirm they are the authors' own creations for this article.

AQ6. Please provide the significance of the bold values in Table 3.

AQ7. The following reference is not included in the reference list: "Educational Testing Service, 2024". Please include the citations in the list with full publication detail.

AQ8. Please provide the missing volume and page range for reference "Amiri et al. (2020)".

AQ9. Please note that the reference "Andéol et al. (2017)" is not cited in the text. Please provide an in-text citation for the reference.

AQ10. Please provide editor for reference "Van Hedger & Johnsrude (2022)".

AQ11. Please check and confirm whether the inserted opening page for reference "Wisniewski et al. (2021)" is correct.

Comparing Speech Intelligibility and Listening Effort Between Native and Non-Native Languages: Application to French Listeners

Caroline Hamery,[1] Vincent Isnard,[2] Marylou Lemouton,[2] Guillaume Andéol,[2] and Sébastien Scannella[1]

AQ1

AQ2 **Objectives:** In a complex acoustic environment, speech intelligibility and listening effort vary depending on both listener characteristics and environmental conditions. A key factor is the language of the speech stimuli, which can influence measured outcomes, particularly when the corpus language differs from the listener's native language. This study compared speech intelligibility and subjective listening effort among 51 French listeners using both the original English and the French version of the coordinate response measure.

Design: To assess performance across different auditory conditions, participants completed tasks in two scenarios: speech-in-noise, to evaluate energetic masking; and speech-in-speech, to assess informational masking.

Results: Participants showed significantly better intelligibility and reduced listening effort in French compared with English. In both conditions, the language effect increased at more challenging masking levels. In the speech-in-speech condition, however, a floor effect emerged beyond a certain threshold. The interaction between language and task difficulty highlights the importance of using a native language corpus to reduce experimental bias.

AQ3

Conclusions: Overall, measuring speech intelligibility in a listener's native language reduces the confounding effects of language proficiency, enabling a more accurate assessment of the effects of listening conditions, intelligibility, and effort.

Key words: Coordinate response measure, Listening effort, Speech-in-noise, Speech-in-speech, Speech intelligibility.

(Ear & Hearing XXX;XX;00–00)

INTRODUCTION

Speech intelligibility tasks are widely used to assess hearing disorders and, more generally, the auditory abilities of participants in challenging listening situations (DiNino et al. 2022). Comparative analyses across experimental and application contexts rely on tasks that are configurable, particularly with respect to participant characteristics and task parameters (e.g., language, spatial distribution of attention), as well as the type of masking (e.g., energetic or informational). Among the tools used to study speech intelligibility, the coordinate response measure (CRM) corpus (Bolia et al. 2000) is widely used, due to its simplicity and adaptability (Brungart 2001a; Mesgarani & Chang 2012; Wisniewski et al. 2021). Like other intelligibility tasks, it enables researchers to assess both speech-in-noise (SIN) and speech-in-speech (SIS) comprehension. In these scenarios, intelligibility can be modulated by adjusting the relative level of the target signal to that of the masker (e.g., background noise or additional talkers). Specifically, the signal to noise ratio (SNR) is used in SIN tests, while the target-to-masker ratio (TMR) is used in SIS tests.

The CRM corpus has been used in both energetic and informational masking paradigms (Brungart 2001b). These phenomena are studied using two simultaneous audio streams that may interfere with each other. The masking effect occurs on at least two levels. First, energetic masking happens when two competing streams of interest overlap in the periphery of the auditory system (Durlach et al. 2003). Second, informational masking refers to the residual interference that remains after accounting for energetic masking (Cooke et al. 2008). This form of masking interacts with energetic masking and is worsened when the streams are semantically similar, as this creates interference that increases the listener's cognitive load during processing (Johnsrude & Rodd 2016).

However, a listener's capacity to understand speech depends not only on their sensory acuity, but also on cognitive factors such as selective attention to a target talker or foreign language proficiency. These factors interact and require the listener to allocate mental resources to understand the target speaker, a demand that has been described in the literature as listening effort (Pichora-Fuller et al. 2016; Peelle 2018). This is particularly relevant in military contexts, where personnel often encounter challenging listening conditions. Ambient interfering sounds, such as ongoing radio communications or engine noise, can create substantial acoustic challenges. The consequence can be increased workload, fatigue, and decreased intelligibility of crucial messages, thereby jeopardizing mission outcomes and individual safety. In the context of high-risk operations, mission success can be compromised if speech intelligibility falls below a critical threshold, even if other sensory information (e.g., visual) is available (Mentel et al. 2013). This observation highlights the importance of improving communication systems and minimizing auditory interference. Therefore, it is vital to maintain speech intelligibility while minimizing listening effort in challenging listening conditions (Tepe et al. 2017).

Furthermore, in complex auditory situations, it becomes even more challenging to understand a second language compared with one's native language (Rogers et al. 2006; Cooke et al. 2008;Warzybok et al. 2015). While this phenomenon can be investigated experimentally, using second-language stimuli to assess speech intelligibility may introduce measurable and avoidable biases. In particular, the original English-language

[1]Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, Toulouse, France; and [2]Institut de Recherche Biomédicale des Armées, Brétigny-sur-Orge, France.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com).

<zdoi; 10.1097/AUD.0000000000001738>

version of the CRM corpus is predominantly used to assess both native and non-native English listeners. This population heterogeneity can result in a wide variation in comprehension abilities, potentially introducing performance bias and undermining reproducibility. For instance, research has reported that a noisy environment affects non-native listeners more (Mattys et al. 2012), and increases their listening effort in SIS tasks (Cooke et al. 2008). In addition, interference between two English audio streams may impair speech intelligibility differently compared with interference between streams in the listeners' native language.

In the present study, we hypothesized that the cognitive load associated with speech intelligibility increases when listening in a second language, thereby increasing the listening effort exerted by the listener. The aim was to quantify the impact of using non-native language stimuli to assess bias in experiments designed to evaluate multi-talker listening situations.

The CRM corpus enables the measurement of a dependent variable, specifically the percentage of correct responses, which is influenced by independent variables such as TMR or SNR. When a non-native language corpus is used, language proficiency is another cofactor. Several versions of the CRM corpus have been developed in languages such as Spanish (Lelo de Larrea-Mancera et al. 2023), Dutch (Nagels et al. 2021), Persian (Amiri et al. 2020), British English (Kitterick et al. 2010; Semeraro et al. 2017), Kannada (Rachana & Neelamegarajan 2024), and Mandarin (Wang et al. 2019). A French version was also recently developed by Isnard et al. (2024).

These localized language versions of the CRM corpus make it possible to eliminate language as a confounding factor and enable a more accurate comparison of the core independent variables. Alternatively, language can be included as an experimental variable in assessments of the additional listening effort associated with processing speech in a non-native language. To the best of our knowledge, however, the impact of CRM language (native versus non-native) on both speech intelligibility and listening effort has not yet been assessed. Therefore, in this paper, we aim to determine the impact of using English and French versions of the CRM corpus on speech intelligibility and listening effort in native French speakers in both SIN and SIS conditions.

MATERIALS AND METHODS

Participants

Fifty-one participants were recruited for this experiment (29 women, 1 other, mean age 26.4 ± 5.4 years). All had normal hearing, as confirmed by pure-tone audiometry using an Elios clinical audiometer (Echodia, Le Mazet-Saint-Voy, France). Hearing thresholds were assessed at the following frequencies: 0.25, 0.5, 1, 2, 4, 6, 8 kHz (hearing level ≤20 dB HL; mean hearing level 0.3 ± 6.75 dB HL) and 12.5 kHz (hearing level ≤20 dB HL; mean hearing level 4.9 ± 13.6 dB HL).

To be eligible, participants had to meet the following criteria: no known hearing impairments, aged between 18 and 40 years, native speakers of French, no uncorrected visual impairments, and adequate English comprehension, subjectively self-assessed using the Language Experience and Proficiency
T1 Questionnaire (LEAP-Q; Marian et al. 2007) (Table 1). Participants received €30 in financial compensation for their involvement in the study.

Stimuli

Stimuli were retrieved from both English and French versions of the CRM corpus. In the original English form, a typical sentence follows a fixed structure: "Ready [call sign], go to [color] [digit] now," where call sign, color, and digit are selected from a predefined word list (Bolia et al. 2000)
(Table 2). The corpus includes eight call signs, four colors, T2
and eight digits, resulting in 256 combinations. These were recorded by eight talkers (4 men and 4 women), making a total of 2048 English-language stimuli. The French-language version of the CRM has been designed to mirror, as closely as possible, the English-language version (Isnard et al. 2024). The sentence structure remains almost identical: "*Prêt* [call sign], *va au point* [color] [number] *go*" with equivalent lexical variables (Table 2). As with the English-language version, eight French talkers (four men and four women) recorded the 256 sentence combinations, resulting in 2048 French-language stimuli. In both languages, the recorded sentences range in duration from 2 to 3 sec.

The two versions of the CRM corpus were used in three distinct experimental conditions:

- SIN condition: In this condition, a single target talker was presented alongside background noise. The SNR corresponded to the ratio of the sound level of the target talker (signal) to the background (noise). The noise was generated as Gaussian noise, spectrally shaped using a finite impulse response filter to match the spectral profile of sentences in the CRM corpus, as described in Brungart (2001b).
- SIS condition: Two talkers were presented simultaneously, with the TMR corresponding to the difference between the sound level of the target and the masker. Talker pairings consisted of either 1 woman and 1 man, 2 women, 2 men, or the same talker. The two simultaneous sentences always featured different call signs and different color/digit combinations.

TABLE 1. Mean and SD for subjective assessment of English proficiency

	Mean	SD
Oral production	6.5	1.7
Oral comprehension	7	1.7
Reading	7.5	1.5
LEAP-Q mean	7	1.5

Oral production, oral comprehension, and reading were assessed on a scale of 1 to 10. LEAP-Q mean represents the mean of the three subjective values.

TABLE 2. The CRM variables in the English and French versions of the corpus

	English	French
Call signs	Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, Tiger	Alpha, Delta, Charlie, Eagle, Kilo, Oscar, Tango, Whiskey
Colors	Blue, Green, Red, White	Bleu, Rouge, Vert, Jaune
Digits	One, Two, Three, Four, Five, Six, Seven, Eight	Un, Deux, Trois, Quatre, Cinq, Six, Sept, Huit
Sentences	"Ready [*call sign*], go to [*color*] [*digit*] now."	"Prêt [*call sign*], va au point [*color*] [*digit*] go."

CRM, coordinate response measure.

- Control condition: A target talker was presented without any masking signal (i.e., in silence), in both languages, to confirm that the participant was able to understand speech clearly in the absence of interference.

Apparatus

Participants were seated in a quiet experimental room, facing a screen that displayed the response button options used in the experimental tasks. Graphical user interfaces were developed in MATLAB R2021a (MathWorks). First, a matrix of 32 buttons was displayed. Buttons were organized into four rows (one for each color) and eight columns (one for each digit) for
F1 the speech intelligibility task (cf. Fig. 1A). Second, subjective listening effort was rated using the Categorical Listening Effort Scaling method (Luts et al. 2010), which quantifies perceived effort in Effort Scale Categorical Units (ESCU) (Rennies et al. 2014). This scale is composed of 7 primary listening effort levels and 6 intermediate levels, making a total of 13 possible answers ranging from "No effort" (score = 1) to "Extreme effort" (score = 13). An additional level, "Only noise" (score = 14), was also included. The scale was presented in French (Lanzilotti 2021) (cf. Fig. 1B).

Auditory stimuli were presented diotically via Beyerdynamic DT-770 headphones, connected to an RME Fireface UCX sound card. Stimuli were delivered at approximately 55 dB SPL, with a sampling rate of 44.1 kHz.

Procedure

After providing informed consent, participants completed pure-tone audiometry. They then completed a demographic questionnaire and the LEAP-Q (Marian et al. 2007). The LEAP-Q enabled us to assess each participant's English comprehension based on measures of language exposure, experience, and proficiency (Table 1). Participants subjectively rated their English proficiency on a scale ranging from 1 to 10 for oral, listening, and reading skills. The average of these three ratings was used as a measure of their subjective proficiency.

Participants then completed the speech intelligibility task. In each experimental block, they were asked to indicate the color and digit associated with the call sign "*Baron*" (in English sentences) or "*Delta*" (in French sentences), by clicking the corresponding button. There was no time limit for responses. The order of presentation of the experimental conditions was counterbalanced, using either the English or French version of the CRM, and either the SIN or the SIS masking condition, according to a predefined Latin square design. For the SIN condition, nine SNR levels were tested (ranging from −18 dB to +6 dB, in 3 dB steps). For the SIS condition, nine TMR levels were used (ranging from −12 dB to +12 dB, also in 3 dB steps).

For each masking level, participants completed 32 trials, representing all possible combinations of 4 colors × 8 digits, selected from recordings by the 8 different talkers. In each trial, the talker(s) were randomly chosen to minimize talker-specific effects.

Last, after each block, participants were asked to rate their subjective listening effort. A control block, consisting of 32 trials, was presented after both SIS and SIN conditions, in both French and English versions. Here, the aim was to ensure that participants clearly understood the stimuli under neutral (no masker) conditions. Each participant completed a total of 38 blocks, each containing 32 trials: 18 SIS blocks (32 trials × 9 TMR levels × 2 languages) + 18 SIN blocks (32 trials × 9 SNR levels × 2 languages) + 2 control blocks (32 unmasked trials × 2 languages). Each experimental block lasted approximately 2 to 3 min, and participants were allowed to take short breaks between blocks. The full experimental session lasted approximately 2 h.

Data Analyses

The percentage of correct responses was calculated as the number of trials in which both the color and digit were correctly identified, divided by the total number of trials in that experimental condition (32 trials in each block). Following current recommendations for speech analyses, scores were normalized using the rationalized arcsine unit (RAU) transform in each condition (Studebaker 1985).

Statistical analyses were carried out using MATLAB and R (version 4.0.3). For both SIN and SIS conditions, a two-way analysis of variance (ANOVA) with Greenhouse–Geisser correction for sphericity was conducted to examine the main effect of CRM language (English versus French) and masking level (9 SNRs or 9 TMRs), as well as their interaction. The significance threshold was set at $p < 0.05$. Tukey's Honestly Significant Difference (HSD) post hoc test was applied for pairwise comparisons when the effect was significant. In the following figures, results are presented as the percentage of correct responses for clarity; however, all statistical analyses were conducted on RAU-transformed data.

Ethics Approval

This study was approved by the local ethics committee (IRB Number 2023 647). All participants provided written informed consent before data collection.

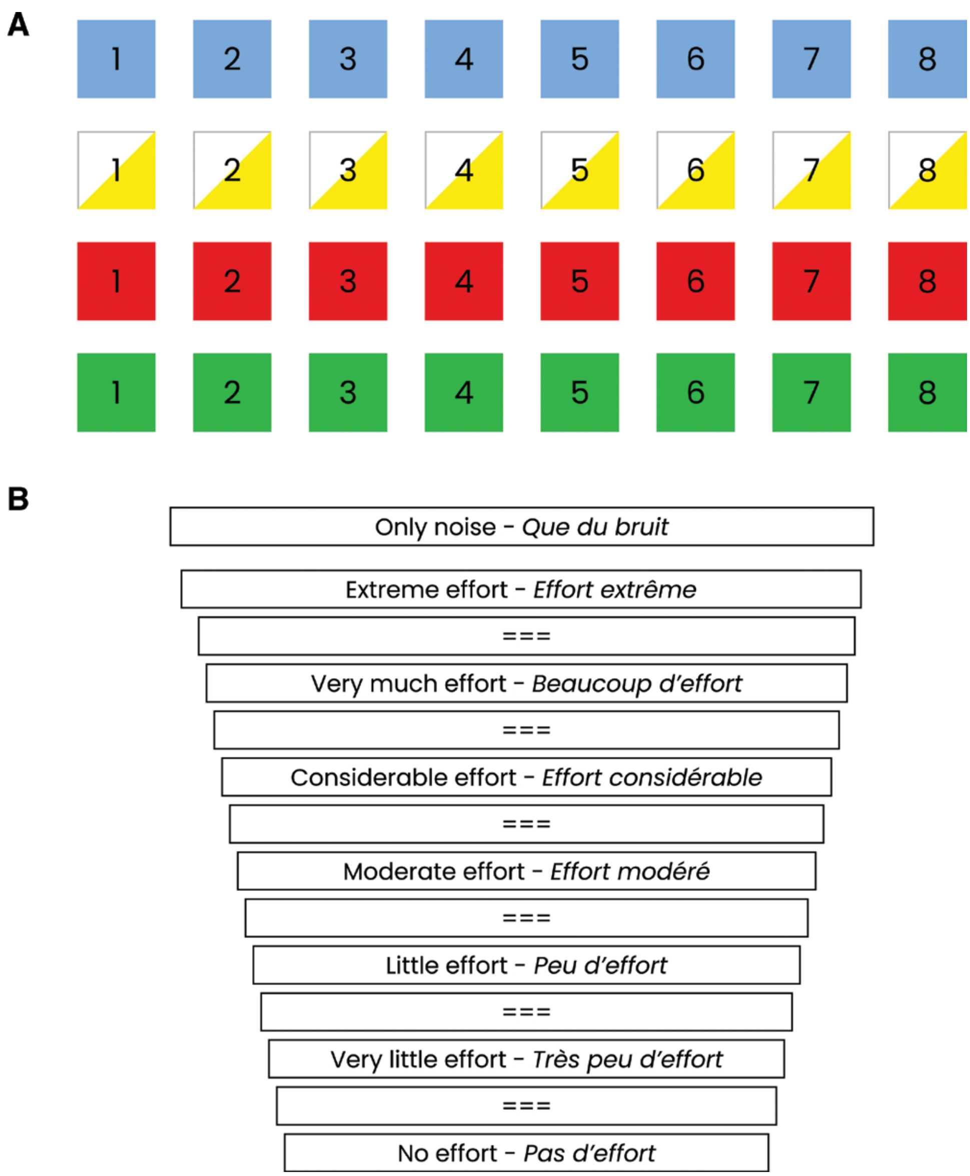
RESULTS

Speech-in-Noise

Speech Intelligibility • In the SIN condition, the ANOVA showed a significant main effect of SNR [$F(8,400) = 2747.81$,
$p < 0.001$, $\eta_p^2 = 0.982$; see Fig. 2A and Table 3]. A significant F2
main effect of CRM language was also observed [$F(1,50) =$ T3
1199.77, $p < 0.001$, $\eta_p^2 = 0.960$], with better intelligibility in French compared with English. In addition, a significant interaction between language and SNR level was found [$F(8,400) = 115.77$, $p < 0.001$, $\eta_p^2 = 0.698$]. Tukey's HSD post hoc test showed that participants performed significantly better in French than in English for negative and neutral SNR levels (i.e., when the masker level was equal to, or higher than the target level), except at the lowest SNR (−18 dB).

Listening Effort • An analysis of ESCU responses after each of the 9 SIN blocks showed a significant main effect of lan-
guage [$F(1,50) = 262.81$, $p < 0.001$, $\eta_p^2 = 0.840$; see Fig. 3A and F3
Table 3], with participants reporting greater subjective listening effort in English than in French. In addition, there was a significant main effect of SNR level on subjective listening effort [$F(8,400) = 844.52$, $p < 0.001$, $\eta_p^2 = 0.944$]; specifically, subjective effort increased as SNR decreased. Furthermore, a significant interaction was observed between language and SNR level [$F(8,400) = 22.33$, $p < 0.001$, $\eta_p^2 = 0.309$]. Tukey's HSD post hoc test showed

LWW 09/20/25 4 Color Fig(s): F1-4 18:26 Art: EANDH-D-24-00601



AQ4 Fig. 1 A, The CRM response matrix. In the English condition, the buttons in the second row were colored white, while in the French condition, they were colored yellow ("*Jaune*" color). B, The Effort Scale Categorical Unit in its English and French versions. Only the French version was used in this study. CRM indicates coordinate response measure. AQ5

that participants reported significantly greater listening effort in English compared with French across all SNR levels, except in the two most adverse conditions (−18 and −15 dB SNR).

Speech-in-Speech

Speech Intelligibility • In the SIS condition, the two-way ANOVA showed a significant main effect of TMR level [$F(8,210) = 268.173$, $p < 0.001$, $\eta_p^2 = 0.843$; see Fig. 2B and Table 2]. Furthermore, a significant main effect of language was observed [$F(1,50) = 270.717$, $p < 0.001$, $\eta_p^2 = 0.844$], with better intelligibility in French compared with English. A significant interaction between language and TMR level was also found [$F(8,400) = 29.531$, $p < 0.001$, $\eta_p^2 = 0.371$]. Post hoc tests revealed that this interaction was due to participants performing better in French than in English at negative and neutral TMR levels (i.e., when the masker was equal to or louder than the target signal).

Listening Effort • A two-way ANOVA of ESCU ratings after each of the nine SIS blocks showed a significant main effect of language [$F(1,50) = 25.71$, $p < 0.001$, $\eta_p^2 = 0.340$; see Fig. 3B and Table 3], with participants reporting greater subjective effort in English compared with French. In addition, a significant main effect of TMR level on subjective listening effort was observed [$F(8,400) = 259.58$, $p < 0.001$, $\eta_p^2 = 0.838$]. A significant interaction between CRM language and TMR level was found [$F(8,400) = 8.13$, $p < 0.001$, $\eta_p^2 = 0.140$]. Tukey's HSD post hoc test showed that participants reported significantly greater listening effort in English compared with French across nearly all TMR levels, except in the two most adverse conditions (−12 and −6 dB TMR).

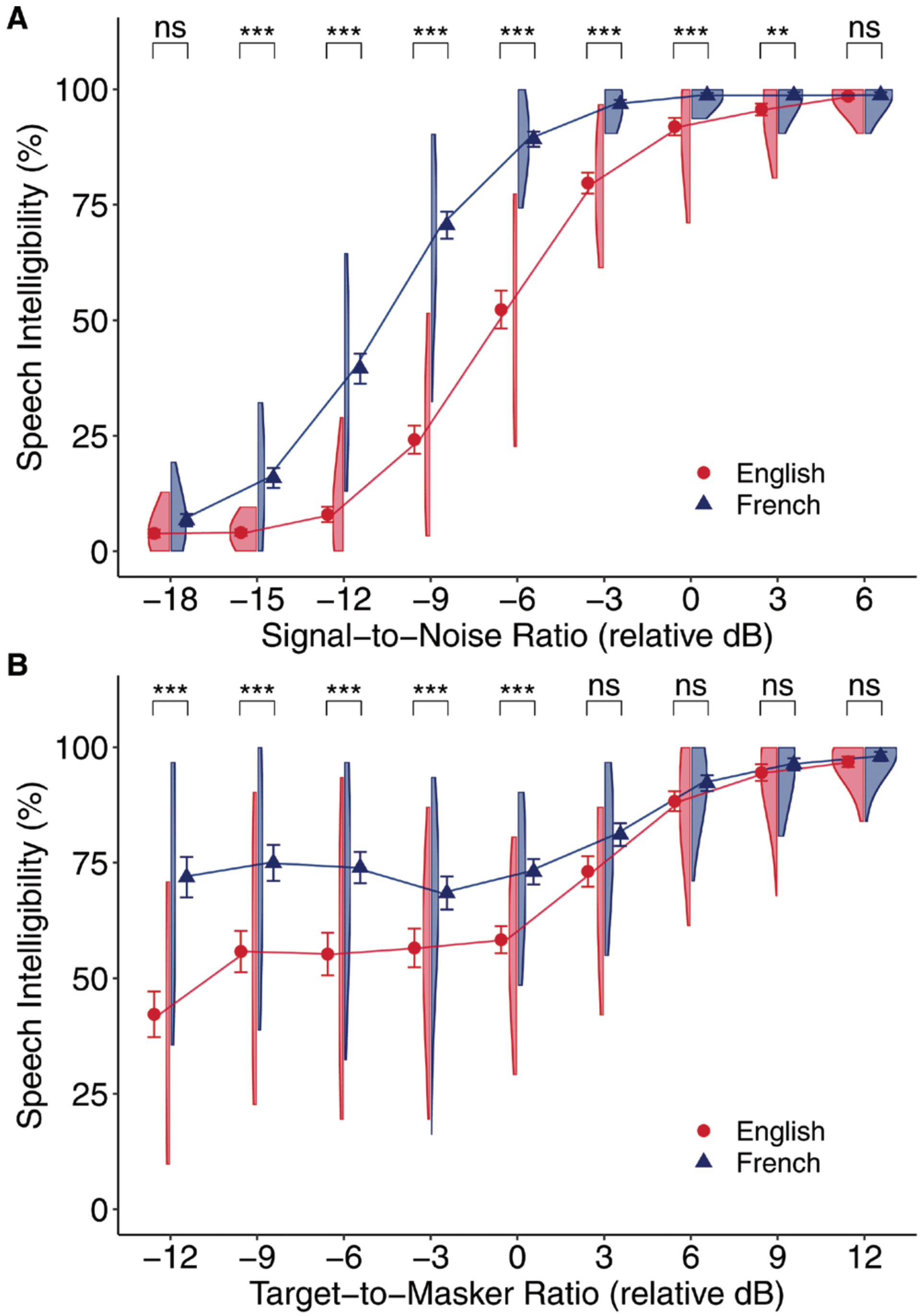


Fig. 2. Mean speech intelligibility (percentage correct). A, For each SNR level (−18 to +6 dB SNR; in 3 dB steps). B, For each TMR level (−12 to +9 dB TMR; in 3 dB steps). Vertical bars indicate SE. Violin plots represent response distribution density. Statistical comparisons were performed on RAU-transformed data. Post hoc significance: **p* < 0.05, ***p* < 0.01, ****p* < 0.001, ns indicates not significant; RAU, rationalized arcsine unit; SNR, signal to noise ratio; TMR, target-to-masker ratio.

Type of Response • To investigate whether incorrect responses in the SIS condition were more influenced by masker interference than random guessing, we investigated participants' responses with respect to digit- and color-related errors. In the majority of incorrect trials, participants preferentially chose information corresponding to the masker rather than making a random decision. This was observed for the color, the digit only, or both. The distribution of response types is illustrated in Figure 4. F4

English Proficiency

Subjective English proficiency was assessed using the LEAP-Q. Three variables were selected: oral production, oral comprehension, and reading. Each variable was rated on a scale

TABLE 3. Main ANOVA results for speech intelligibility (CRM) and listening effort (ESCU) in SIS and SIN conditions

		Speech Intelligibility (CRM)				Listening Effort (ESCU)			
		Df	F	η_p^2	*p*	Df	*F*	η_p^2	*p*
SIS	SNR	8	2600.42	0.981	**<0.001**	8	844.52	0.944	**<0.001**
	Language	1	1048.25	0.954	**<0.001**	1	262.81	0.840	**<0.001**
	Language:SNR	8	91.66	0.647	**<0.001**	8	22.33	0.309	**<0.001**
SIN	TMR	8	293.17	0.854	**<0.001**	1	259.58	0.838	**<0.001**
	Language	1	246.90	0.832	**<0.001**	8	25.71	0.340	**<0.001**
	Language:TMR	8	23.98	0.324	**<0.001**	1	8.13	0.140	**<0.001**

CRM, coordinate response measure; Df, degrees of freedom; ESCU, Effort Scale Categorical Units; F, F test value, η_p^2, effect size; SIN, speech-in-noise; SIS, speech-in-speech.

AQ6

from 1 to 10, with 10 representing perfect proficiency. The mean of these three ratings was used as the overall proficiency score. The mean LEAP-Q score across participants was 7.0 ± 1.5 (SD). Results are shown in Table 1.

This measure confirmed a minimum level of English proficiency in the same way that audiometric testing ensured that participants had no hearing loss.

DISCUSSION

The goal of the present study was to evaluate the effects of language (native versus non-native) on speech intelligibility and listening effort. English and French versions of the CRM corpus were used with native French listeners under two conditions: (1) SIN, in which a single talker was masked by stationary speech-shaped noise; and (2) SIS, involving two simultaneous talkers. Listening effort was assessed in each condition using the Categorical Listening Effort Scale method (Luts et al. 2010; Rennies et al. 2014).

Speech-in-Noise • As expected, in the SIN condition, intelligibility gradually declined as SNR decreased. The main effect of SNR is aligned with previous findings reported in the literature (Brungart 2001a, 2001b). We also found that intelligibility in the presence of background noise was influenced by the language. As expected, French participants performed better when listening in French compared with English. This suggests that the louder the relative masking noise, the more difficult it is for the listener to separate the target from the noise, and that this challenge is greater when the speech is in a non-native language.

These intuitive results provide useful reference values for SIN tests. With the exception of the most adverse condition (−18 dB SNR), where a floor effect was observed, and the most favorable conditions (+3 and +6 dB SNR), where intelligibility reached its maximum, participants consistently performed better in French than in English. This pattern suggests a language-related effect that interacts with the level of energetic masking.

The significant interaction between language and SNR, however, is of particular interest. Our results suggest that energetic masking influences speech processing differently in native and non-native languages. In French, intelligibility differed significantly between the two most adverse SNR levels (−18 and −15 dB SNR; see Figure 1A in Supplemental Digital Content, https://links.lww.com/EANDH/B750, for Tukey's post hoc results), while this was not the case in English. In addition, a ceiling effect appeared at lower SNR levels in French compared with English (no significant difference between −3 and 0 dB SNR; see Figure 1A in Supplemental Digital Content, https://links.lww.com/EANDH/B750), showing that participants were able to cope better with louder background noise in their native language.

The absence of a language effect in the most adverse condition can be explained by the high level of masking noise. In this condition, whatever the language, masking noise was loud enough to prevent any understanding of the target speech, as demonstrated by very low intelligibility scores. ESCU ratings for this SNR level provide further support for the disengagement or withdrawal of participants during this block; most participants selected the "only noise" level.

Speech-in-Speech • In the SIS condition, the results confirmed our first hypothesis regarding a significant effect of task difficulty (i.e., manipulated by TMR levels) on speech intelligibility. Intelligibility was better at positive TMRs, where the target voice was louder than the masker. This main effect, which is consistent with previous studies (Brungart 2001b; Thompson et al. 2015), shows that intelligibility decreases as the masker becomes louder relative to the target. Earlier work has shown that informational masking in competing speech streams interferes with speech processing and makes segregation more difficult (Johnsrude & Rodd 2016).

In SIS tasks, masking is both energetic and informational. Evidence for informational masking can be found in the nature of participants' errors. A closer examination of response types highlighted that incorrect answers were predominantly biased toward the competing talker. It is known that the more similar the competing voices, the more difficult it is to segregate them (Brouwer et al. 2012). Thus, the errors made by participants could be related to both the TMR and/or the acoustic characteristics of the voices (Brungart 2001b). In addition, as in the SIN condition, the auditory stream from the masker talker introduces a degree of energetic masking. However, the result reported here, namely that errors align with the masker talker, strongly suggests that the dominant form of interference was informational in the SIS condition.

The effect of task difficulty on intelligibility was observed in both the French and English versions of the CRM corpus. However, as in the SIN condition, the effect was more pronounced in English. Participants were better able to segregate the target talker when listening in French compared with English. This is probably because they were all native French listeners. Earlier work has demonstrated that knowledge, familiarity, and expertise minimize informational masking in multi-talker scenarios (Johnsrude et al. 2013). This significant primary effect of language is underlined by large effect sizes in both conditions (SIN: $\eta_p^2 = 0.982$; SIS: $\eta_p^2 = 0.843$).

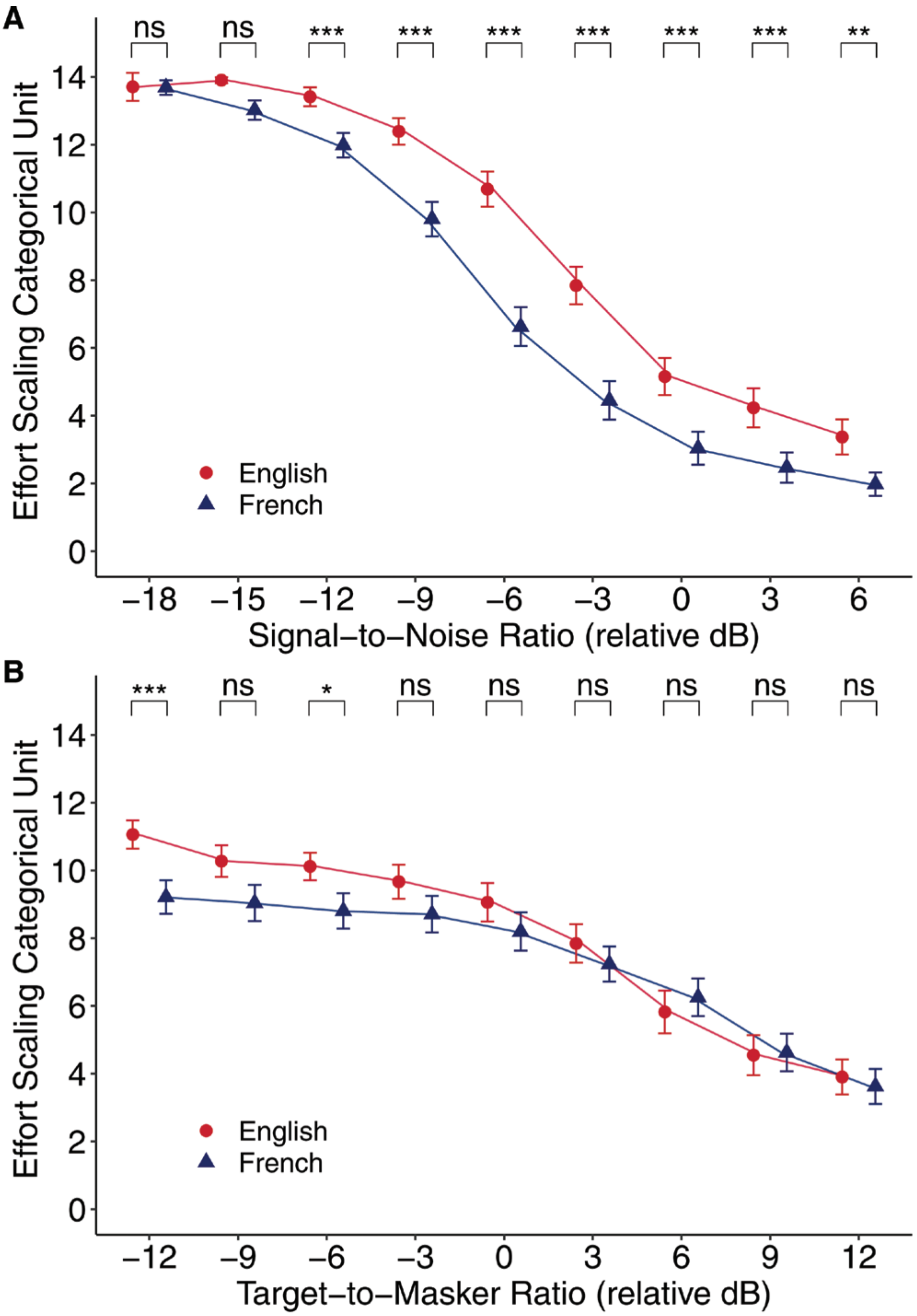


Fig. 3. Mean effort scaling categorical unit responses. A, For each SNR level (−18 to +6 dB SNR; in 3 dB steps). B, For each TMR level (−12 to +9 dB TMR; in 3 dB steps). The scale ranges from 0 (no effort) to 14 (only noise). Vertical bars indicate SE. Statistical comparisons were performed on RAU-transformed data. Post hoc significance: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, ns indicates not significant. RAU, rationalized arcsine unit; SNR, signal to noise ratio; TMR, target-to-masker ratio.

Overall, mean intelligibility scores for English and French ranged from 2 to 28 RAU, with higher scores in French. In addition, performance in the French language tended to stabilize around 70 RAU for the adverse condition, whereas it rapidly decreased in the English language, reaching a minimum of 40 RAU at −12 dB TMR. This marked drop shows how, in a non-native language, informational masking not only affects intelligibility at a lower TMR, but also that the effect is stronger compared with a native language.

Because the energetic masking produced by the competing talker is similar in both languages, the difference in performance is most likely to be driven by increased informational masking in the non-native language. This finding supports the idea that informational masking has a stronger impact when listeners process speech in a non-native language, particularly at the lowest TMRs (Van Engen & Bradlow 2007; Lecumberri et al. 2010).

LWW | 09/20/25 | 4 Color Fig(s): F1-4 | 18:26 | Art: EANDH-D-24-00601

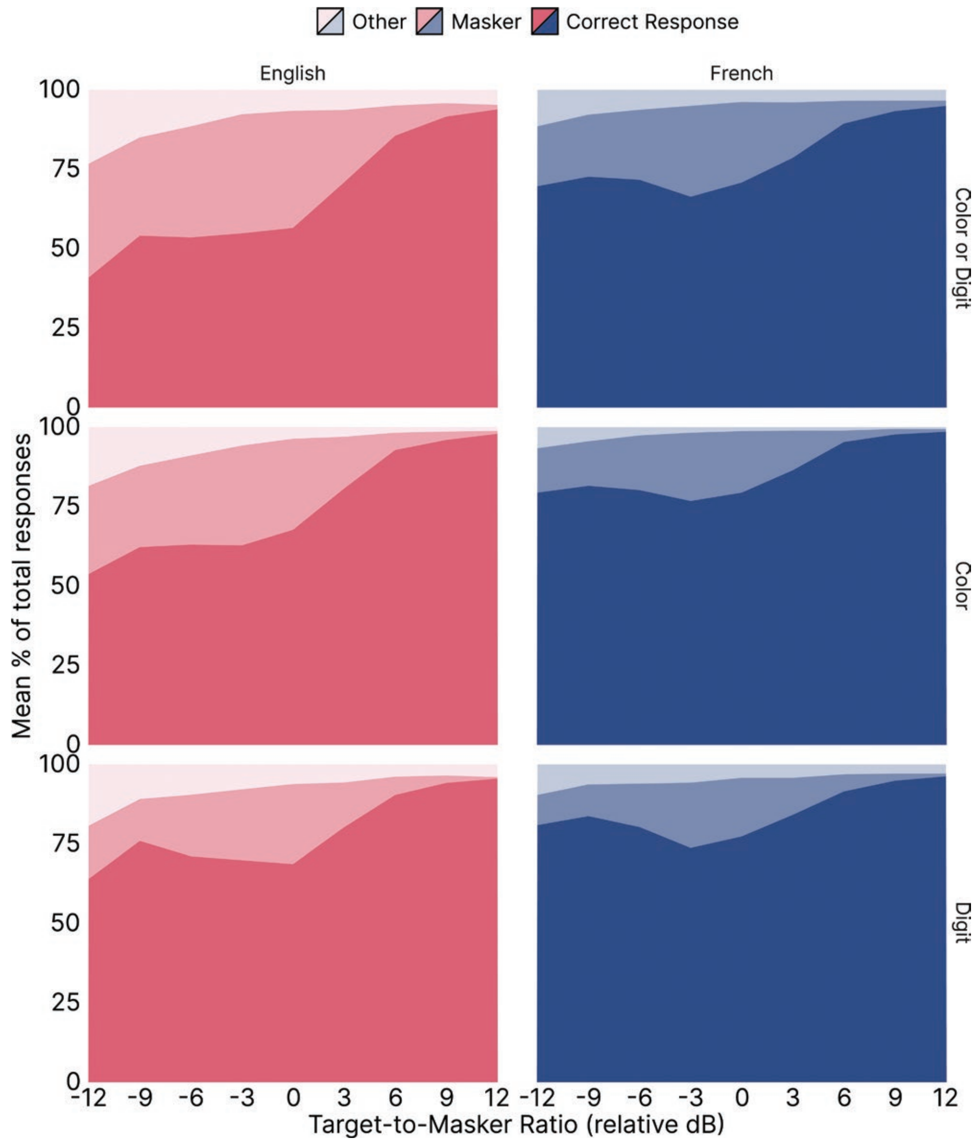


Fig. 4. Mean percentage of response type for English and French CRM in the SIS condition. Correct responses, masker responses, or other responses for color, digit, or both, for each TMR level. CRM indicates coordinate response measure; SIS, speech-in-speech; TMR, target-to-masker ratio

The primary outcome of this study, however, reveals an interaction between the CRM language and the TMR level in the SIS condition. Under adverse conditions (i.e., when the target talker was at the same or lower auditory level than the masker), French listeners were better at segregating the target talker in French compared with English, while intelligibility did not significantly differ in favorable conditions. French listeners performed better in the French SIS condition in the range +3 to −12 dB TMR, and their subjective listening effort was rated as significantly higher in English compared with French.

While, in general, the effects of energetic masking decrease monotonically with increasing SNR (Brungart 2001b; Thompson et al. 2015), the presence of informational masking introduces additional variability and engages more complex cognitive processes (Mattys et al. 2012). Examination of individual intelligibility curves revealed distinct performance patterns, suggesting that participants may have different segregation and inhibition strategies. Some showed a monotonic relation between TMR and intelligibility (i.e., the lower the TMR, the lower the intelligibility, as observed in the SIN condition), while for others, intelligibility was minimal at 0 dB TMR, and performance was equal or better for the most adverse TMRs.

The results showed that 30 participants performed equally well or better in the range −12 to 0 dB TMR in the French condition, whereas this pattern was only observed in 12 participants in the English condition (see Figure 2 in Supplemental Digital Content, https://links.lww.com/EANDH/B751). As reported by other authors (Andeol et al. 2011; Thompson et al. 2015; Lanzilotti et al. 2022), these results can be explained by the listener's ability to use the cue of a softer voice to segregate the target talker from the masker. While both native and non-native listeners tend to adopt similar strategies when actively listening in complex situations, beyond a certain level, non-native listeners are more adversely affected by environmental distortion and complexity (Bradlow & Alexander 2007; Brouwer et al. 2012).

In our study, we argue that an additional layer of complexity emerged: listeners seemed to be more adept at leveraging these "negative level cues" in their native language.

It is also important to note that performance significantly improved between −12 and −9 dB TMR in English, whereas no such improvement was observed in French. In this TMR range, the psychometric function displayed floor and ceiling effects, demonstrated by the absence of significant differences in adjacent levels (from −12 to 0 dB, and from +6 to +12 dB). Similarly, we observed significant differences between 0 and +3 dB TMR and between +3 and +6 dB TMR in English. The floor effect, however, appeared to persist beyond the initial decline, as intelligibility remained significantly different between −9 and −12 dB TMR (see Results in Supplemental Digital Content, https://links.lww.com/EANDH/B750, and https://links.lww.com/EANDH/B751).

We suggest that this result may be due to listener disengagement under the most adverse conditions in English. Previous research has shown that disengagement can occur when task difficulty is thought to be beyond a perceived level of achievability (Hopstaken et al. 2015). This result also illustrates the fall in intelligibility in a non-native language compared with a native one. We suggest that the same effect could have been observed in French if we had tested more adverse TMR levels (see Figure 1B in Supplemental Digital Content, https://links.lww.com/EANDH/B750).

The Effect of Language on Energetic and Informational Masking • In SIN and SIS conditions, masking affects intelligibility at two distinct, but simultaneous processing levels. In SIN conditions, pure energetic interference disturbs speech processing at the peripheral level. In SIS conditions, complementary semantic interference makes segregation and comprehension processes more difficult.

Nevertheless, our results demonstrate that intelligibility under adverse conditions becomes more challenging when speech is presented in a non-native language compared with a native language, whatever the type of interference. Moreover, our findings suggest that processing speech stimuli presented in a non-native language under adverse conditions increases cognitive load. This cost is observable both in terms of speech intelligibility degradation and subjective listening effort.

Segregation between the target and masker was shown to be easier for native French listeners when the speech was presented in French rather than in English. Naturally, native listeners are experts, both at the semantic and linguistic levels, as they are more exposed to this language. They may also be more likely to associate French voices with familiar characteristics, compared with English voices. Furthermore, as it is known that familiar voices are more easily segregated (Johnsrude et al. 2013), masking in French may have had less impact than masking in English for our participants.

Lecumberri et al. (2010) noted that non-native listeners face the dual challenge of an imperfect signal and imperfect knowledge. Similarly, in SIN scenarios, non-native listeners have consistently shown greater difficulty with intelligibility than native listeners, due to energetic masking (Lecumberri & Cooke 2006; Rogers et al. 2006; Mattys et al. 2012).

In the present study, our participants knew some English (mean LEAP-Q score 7.01/10) but were not native. Other studies have used unfamiliar language to assess the influence of language on intelligibility in complex auditory environments (Rhebergen et al. 2005; Lecumberri & Cooke 2006; Van Engen & Bradlow 2007; Brouwer et al. 2012). For example, Brouwer et al. (2012) showed that an unfamiliar language mask had a relatively small impact on intelligibility compared with a familiar or native language.

The intelligibility differences we observed between languages in adverse conditions are corroborated by our results for subjective listening effort. In both SIS and SIN conditions, participants reported that tasks that used the English version of the CRM corpus were more effortful than those that used the French corpus. The interaction between language, energetic, and informational masking supports the idea that speech segregation requires more cognitive resources in a non-native language—even for simple words. In our study, intelligibility in adverse conditions (e.g., −9 dB SNR or TMR) was, on average, similar in SIN and SIS conditions in French. However, this was not the case for the English corpus. Here, the percentage of correct responses decreased to 20% in the SIS condition, while it stabilized at around 50% in the SIN condition. This difference may be due to additional informational processing in the SIS condition. As Van Hedger and Johnsrude (2022) noted in their review of speech perception in adverse conditions, "different adverse conditions place different demands on cognitive resources."

Based on the results presented here, we suggest that the "knowledge-driven" processes described by Mattys et al. 2012, which listeners rely on to perceive speech in complex auditory situations, are more fragile when operating in a non-native language, and come with a higher cognitive cost (Lecumberri & Cooke 2006; Golestani et al. 2009). This difference is clearly illustrated by the better intelligibility of the French version of the CRM corpus compared with the English version in the most adverse conditions. Last, we suggest that the difference described above is mainly due to the increased cognitive demands involved in comprehending and retrieving information when processing speech in a non-native language in complex conditions.

Furthermore, the observation that intelligibility differences between the two languages are inhomogeneous across masking levels suggests that using a foreign language to assess SIN and SIS processing may be less reliable than using participants' native language. If the observed differences between French and English had been significant and homogeneous across all masking levels, the impact would be more predictable—using an English rather than a native language corpus would simply be associated with lower average performance and increased listening effort. However, the present results indicate that the impact of the use of an English corpus on intelligibility and listening effort is most pronounced under adverse conditions, particularly those involving high levels of masking. This results in heterogeneous costs across TMR and SNR levels. Consequently, we argue that the loss of intelligibility in a non-native language is related to increased listening effort, but this is highly dependent on the complexity of the auditory environment.

Language Proficiency

The purpose of this study was to evaluate the impact of using an English-language compared with a French-language CRM corpus on speech intelligibility and listening effort in a sample of native French listeners. English proficiency was subjectively assessed to ensure that participants understood the stimuli. Although we made no hypotheses about English proficiency levels, it is important to acknowledge that language proficiency significantly affects intelligibility in non-native listeners, especially under complex auditory conditions (Rhebergen et al. 2005; Lecumberri & Cooke 2006; Rogers et al. 2006; Cooke & Lecumberri 2012; Warzybok et al. 2015).

For example, Smiljanić and Bradlow (2011) showed that language proficiency affects intelligibility and noted that more experienced non-native listeners are better able to cope with background noise than less experienced ones. Similarly, Warzybok et al. (2015) demonstrated that both language proficiency and the type of speech task have a strong influence on intelligibility. However, given the restricted set of response alternatives and the simplicity of the corpus language, the CRM task used in our study may be less sensitive to differences in proficiency.

Our study could have been strengthened by including an objective assessment of English proficiency, based on more robust tests. Including language proficiency as an additional factor in our analyses would have provided a more nuanced interpretation of the results. For instance, standardized English AQ7 tests such as the TOEFL (Educational Testing Service 2024) or IELTS (British Council, IDP: IELTS Australia & Cambridge Assessment English 2024) could have been used.

CONCLUSION

This study investigated the influence of native (French) and non-native (English) languages under conditions of energetic and informational masking. The main result is that differences in intelligibility and listening effort between the two languages are not homogeneous across task difficulty but rather depend on the relative masking level.

These results are of interest for the design of further studies and could be used to investigate listening effort in complex auditory situations, particularly when incorporating objective electrophysiological measures. The observed decline in intelligibility in English, compared with French, under the most adverse conditions, supports the hypothesis that using a non-native language may lead to information loss for non-native listeners, even with simple auditory stimuli.

Further studies could extend this work by investigating other foreign language pairs, as this would help to assess the reliability of the findings presented here and explore whether effects may differ depending on the corpus language.

This work provides a foundation for future studies on listening effort using an appropriate corpus. It paves the way for further experiments incorporating physiological measures of listening effort, as a complement to subjective assessments, to gain a more comprehensive understanding of cognitive load in challenging listening conditions.

ACKNOWLEDGMENTS

All authors contributed equally to this work. C.H. experimental design, data acquisition, data analysis, synthesis, writing, proofreading; M.L. experimental design; V.I., G.A., and S.S. experimental design, data analysis, synthesis, writing, proofreading. All authors discussed the results and implications and commented on the manuscript at all stages.

This study has been approved by the local ethics committee: Toulouse University, IRB N°2023647. All participants provided written informed consent before any data collection.

The data that support the findings of this study are openly available at https://osf.io/ye4f7/?viewonly=afa14ede0dc1480eb031e0235bea2b8c.

The authors have no conflicts of interest to disclose.

Address for correspondence: Hamery Caroline, DCAS—Institut Supérieur de l'Aéronautique et de l'Espace - Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, 10 Avenue Marc Pélegrin, Toulouse, 31055, France. E-mail: caroline.hamery@isae-supaero.fr

Received December 9, 2024; accepted August 19, 2025

REFERENCES

Amiri, M., Jarollahi, F., Jalaie, S., Sameni, S. J. (2020). A new speech-in-noise test for measuring informational masking in speech perception among elderly listeners. *Cureus* https://doi.org/10.7759/cureus.7356. AQ8

Andeol, G., Guillaume, A., Micheyl, C., Savel, S., Pellieux, L., Moulin, A. (2011). Auditory efferents facilitate sound localization in noise in humans. *J Neurosci*, *31*, 6759–6763.

Andéol, G., Suied, C., Scannella, S., 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. AQ9

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

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J Acoust Soc Am*, *121*, 2339–2349.

Brouwer, S., Van Engen, K. J., Calandruccio, L., Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *J Acoust Soc Am*, *131*, 1449–1464.

Brungart, D. S. (2001a). Evaluation of speech intelligibility with the coordinate response measure. *J Acoust Soc Am*, *109*, 2276–2279.

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

Cooke, M., Garcia Lecumberri, M. L., Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *J Acoust Soc Am*, *123*, 414–427.

Cooke, M., & Lecumberri, M. L. G. (2012). The intelligibility of Lombard speech for non-native listeners. *J Acoust Soc Am*, *132*, 1120–1129.

DiNino, M., Holt, L. L., 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.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., Kidd, G. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *J Acoust Soc Am*, *114*, 368–379.

Golestani, N., Rosen, S., Scott, S. K. (2009). Native-language benefit for understanding speech-in-noise: The contribution of semantics. *Bilingualism*, *12*, 385–392.

Hopstaken, J. F., Van Der Linden, D., Bakker, A. B., Kompier, M. A. J. (2015). A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology*, *52*, 305–315.

Isnard, V., Chastres, V., Andéol, G. (2024). French version of the coordinate response measure corpus and its validation on a speech-on-speech task. *JASA Express Lett*, *4*, 075203.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., Carlyon, R. P. (2013). Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychol Sci*, *24*, 1995–2004.

Johnsrude, I. S., & Rodd, J. M. (2016). Chapter 40—Factors that increase processing demands when listening to speech. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 491–502). Academic Press.

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

Lanzilotti, C. (2021). *Étude multidimensionnelle de l'effort d'écoute en situations multilocuteurs: Mesures subjectives, comportementales et en* *imagerie optique*. [Phd thesis, Sorbonne Université]. https://tel.archives-ouvertes.fr/tel-03533020.

Lanzilotti, C., Andéol, G., Micheyl, C., 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.

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

Lecumberri, M. L. G., Cooke, M., Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Commun*, *52*, 864–886.

Lelo de Larrea-Mancera, E. S., Solís-Vivanco, R., Sánchez-Jimenez, Y., Coco, L., Gallun, F. J., 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.

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., Spriet, A. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *J Acoust Soc Am*, *127*, 1491–1505.

Marian, V., Blumenfeld, H. K., Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *J Speech Lang Hear Res*, *50*, 940–967.

Mattys, S. L., Davis, M. H., Bradlow, A. R., Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Lang Cognit Process*, *27*, 953–978.

Mentel, K., Ziriax, J., Dachos, J., Salunga, A., Turner, H., Sheffield, B., Brungart, D. (2013). The relationship between speech intelligibility and operational performance in a simulated Naval Command Information Center. *Proc Hum Factors Ergon Soc Annu Meet*, *57*, 1204–1208.

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, *485*, 233–236.

Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., 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.

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear*, *39*, 204–214.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear Hear*, *37*, 5S.

Rachana, D., & Neelamegarajan, D. (2024). Development and content validation of coordinate response measure (CRM) corpus in Kannada for informational masking measurement. *Egypt J Otolaryngol*, *40*, 110.

Rennies, J., Schepker, H., Holube, I., Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *J Acoust Soc Am*, *136*, 2642–2653.

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

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Appl Psycholinguist*, *27*, 465–485.

Semeraro, H. D., Rowan, D., van Besouw, R. M., Allsopp, A. A. (2017). Development and evaluation of the British English coordinate response measure speech-in-noise test as an occupational hearing assessment tool. *Int J Audiol*, *56*, 749–758.

Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *J Acoust Soc Am*, *130*, 4020–4031.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J Speech Lang Hear Res*, *28*, 455–462.

Tepe, V., Smalt, C., Nelson, J., Quatieri, T., Pitts, K. (2017). Hidden hearing injury: The emerging science and military relevance of cochlear synaptopathy. *Mil Med*, *182*, e1785–e1795.

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

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

Van Hedger, S. C., & Johnsrude, I. S. (2022). Speech perception under adverse listening conditions. In *Speech Perception* (pp. 141–171). Springer. AQ10

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

Warzybok, A., Brand, T., Wagener, K. C., Kollmeier, B. (2015). How much does language proficiency by non-native listeners influence speech audiometric tests in noise? *Int J Audiol*, *54*(Suppl 2), 88–99.

Wisniewski, M. G., Zakrzewski, A. C., Bell, D. R., Wheeler, M. (2021). EEG power spectral dynamics associated with listening in adverse conditions. *Psychophysiology*, *58*, e13877. AQ11