Development of a French version of the Modified Rhyme Test

Véronique Zimpfer, Guillaume Andeol, Geoffroy Blanck, Clara Suied, Thibaut
Fux

Development of a French version of the Modified Rhyme Test

Véronique Zimpfer, Guillaume Andéol, Geoffroy Blanck, Clara Suied, and Thibaut Fux

---

ARTICLES YOU MAY BE INTERESTED IN

Dong speakers outperform Mandarin speakers in behavioral pitch discrimination
The Journal of the Acoustical Society of America **147**, EL62 (2020); https://doi.org/10.1121/10.0000604

Ganong effects for frequency may not be robust
The Journal of the Acoustical Society of America **147**, EL37 (2020); https://doi.org/10.1121/10.0000562

Evidence supporting synchrony between two active ears due to interaural coupling
The Journal of the Acoustical Society of America **147**, EL25 (2020); https://doi.org/10.1121/10.0000473

A note on wind velocity and pressure spectra inside compact spherical porous microphone windscreens
The Journal of the Acoustical Society of America **147**, EL43 (2020); https://doi.org/10.1121/10.0000592
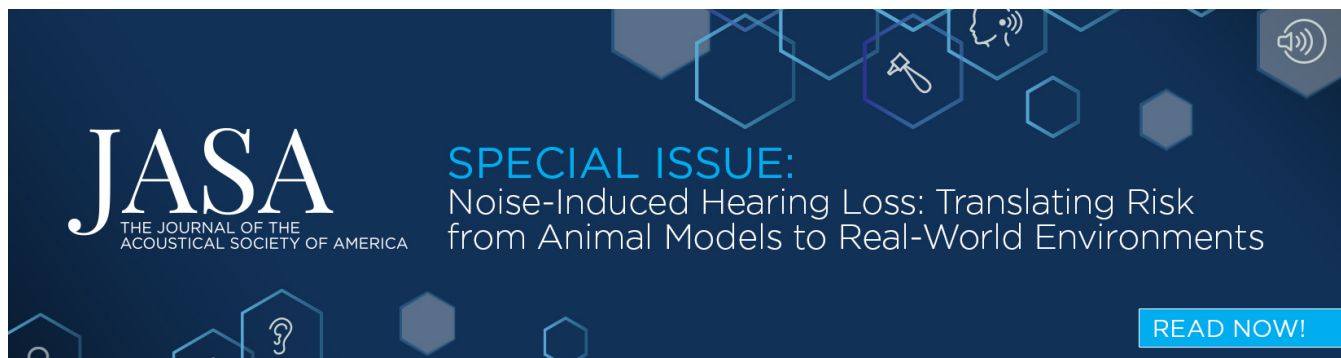
The effects of speech intelligibility on English scientific literature reading in Chinese open-plan offices
The Journal of the Acoustical Society of America **147**, EL1 (2020); https://doi.org/10.1121/10.0000497

Voice fundamental frequency differences and speech recognition with noise and speech maskers in cochlear implant recipients
The Journal of the Acoustical Society of America **147**, EL19 (2020); https://doi.org/10.1121/10.0000499

Development of a French version of the Modified Rhyme Test

Véronique Zimpfer,[1,a] Guillaume Andéol,[2] Geoffroy Blanck,[3] Clara Suied,[2] and Thibaut Fux[2]

[1]*French–German Research Institute of Saint-Louis (ISL), 5 rue du Général Cassagnou, BP 70034, 68 301 Saint-Louis, France*
[2]*Institut de Recherche Biomédicale des Armées (IRBA), 1 place du générale Valérie André, BP 73, 91223 Brétigny-sur-Orge, France*
[3]*Direction Générale de l'Armement Techniques Terrestres (DGA TT), Rocade Est – Echangeur de Guerry 18021 Bourges Cedex, France*
Veronique.zimpfer@isl.eu, guillaume.andeol@def.gouv.fr, geoffroy.blanck@intradef.gouv.fr, clara.suied@def.gouv.fr, thibaut.fux@gmail.com

Abstract: A reproducible method to measure the intelligibility of communication systems is required to assess their efficiency. The current study seeks to develop a French version of the Modified Rhyme Test (MRT) [(House, Williams, Heker, and Kryter (1965). J. Acoust. Soc. Am. **37**, 158–66], an intelligibility test composed of 50 six-word lists, originally developed for military applications and now widely used. An evaluation of the authors' French MRT was carried out, reproducing the original experimental conditions used by House and colleagues. Very similar results were found between the original MRT and the French MRT, validating the latter for further use and dissemination.
© 2020 Acoustical Society of America

1. Introduction

The assessment of speech intelligibility of radio communication systems is crucial to prevent communication errors and, therefore, potential accidents. Two kinds of methods are commonly used: behavioral methods or acoustical measurements methods. When both methods can be applied, their results are highly correlated (Griffin, 1992; Yu *et al.*, 2010; Brammer *et al.*, 2011; Anderson and Kalb, 1987). However, acoustical methods cannot be used with radio communication systems that use bone conduction devices because of the lack of reliable transfer function for these devices (Pellieux *et al.*, 2005). In military environments, bone conduction headphones are widely used because their compatibility with nonlinear earplugs maintains good situation awareness. They also offer the possibility to be combined with bone conduction microphones (Manning *et al.*, 2017). Therefore, behavioral methods are crucial for measuring speech intelligibility in military environments.

The Modified Rhyme Test (MRT) has first been developed to assess speech intelligibility in military communication systems (House *et al.*, 1965). During a MRT task, the listener has to recognize a spoken word among a list of six written possible monosyllabic words, differing in their initial or in their final consonant. These consonant sounds are easily distorted during radio communication. Therefore, the MRT appears particularly relevant in this context. MRT is, indeed, the most popular behavioral method of speech intelligibility in military environments (Letowski and Scharine, 2017; MIL-STD-1474E, 2015). MRT is also widely used in civilian applications (Brandewie and Zahorik, 2011; Brammer *et al.*, 2017; Barker and Coyne, 2017) and has been chosen by the American National Standards Institute (ANSI/ASA S3.2, 1989) as the standard method for assessing speech intelligibility in communication systems.

In addition, the MRT has several specificities that explain its popularity, in particular for the military application. First, it is a closed-set speech test, thus reducing procedural learning. Second, with noisy background, the variation of MRT scores is highly correlated with the variation of the Speech Transmission Index (STI), an acoustical metric (Brammer *et al.*, 2011; Anderson and Kalb, 1987).

Because of those advantages, several adaptations of the MRT in other languages have been suggested in the literature: Turkish (Arıöz and Günel, 2015), Czech (Tihelka and Matoušek, 2004), Spanish (Ball, 2011), and Catalan (Alías and Triviño, 2007).

---

[a]Author to whom correspondence should be addressed.

In the present study, we developed a French version of the original MRT (hereafter called Fr-MRT), and conducted a behavioral assessment of it. To assess the Fr-MRT, we varied, based on the above-mentioned literature, three main parameters: the Signal-to-Noise Ratio (SNR), the talker, and a potential learning effect. We then compared the results obtained for the Fr-MRT with STI results, in order to contrast the behavioral results with an objective measurement of speech intelligibility.

We obtained results with the Fr-MRT that were similar to the original MRT, validating the latter for further use and dissemination. We provide in Table 1 the 50 lists of 6 words composing the Fr-MRT.

2. Creation of the word lists

The MRT (House et al., 1965) consists of 50-word lists of American English monosyllabic words. Each list is composed of 6 consonant-vowel-consonant (CVC) words [with few consonant-vowel (CV) and vowel-consonant (VC) words] with an identical central vowel. For a given list, the six words begin with, or end with, the same consonant. Thus, in each list, only one consonant phoneme differs between the six words. In the original MRT, 21 consonants, one semi-consonant, and 10 vowels were used. This is to be compared with a total of 22 to 24 consonants, 2 semi-consonants, and 15 to 19 in American English (Bizzocchi, 2017). The French language is composed of 20 consonants (with 3 semi vowels) and 14 vowels (Calliope and Fant, 1989).

The word lists of the Fr-MRT were chosen using the "Lexique3" French database (New et al., 2004), which provide for each word the occurrence frequency and a measure of the word familiarity. Using the database, monosyllabic CVC words with identical final VC or initial CV were automatically gathered. We obtained more than 120 different lists. In order to select the least bias as possible 50 lists, we first removed the lists that did not have at least 6 words, and second, we ranked the lists on the basis of the occurrence frequency statistics, so that we kept only words as familiar as possible. In order to reduce the number of lists and the number of words within each list, we tried to have a similar frequency of consonants occurrence in the lists [see Fig. 1(A)] as in natural French spoken language (Combescure, 1981). Finally, as in the original MRT, half of the 50 lists were contrasting the initial consonant and the second half the final consonant. The 4 final lists (lists 46–50; with the initial consonant varying) were designed to target specificities of the French language: they are composed of monosyllabic non-CVCs, with 2 semi-vowels. The lists of the Fr-MRT are presented in Table 1. Figure 1(B) compares the word occurrence frequency in each corpus language for Fr-MRT (New et al., 2004) and original MRT (OANC, 2015).

3. Validation of the lists

3.1 Participants

Twenty-six normal hearing listeners (8 females and 18 males) took part in the present study (average of 26 years old). All participants were French-native speakers. All had audiometric threshold below or equal to 20 dB hearing level for all octave frequencies from 125 to 8000 Hz. They all gave written informed consent to participate in this experiment. The study was approved by the ethics committee of Occitany Region Regional Health Agency (IDRCB 2017-A00859–44). For practical reasons, listeners were recruited and tested in two sites (the French forces biomedical research institute, in the Paris region and the French-German research center of Saint Louis). The procedure and the materials were identical in both sites.

3.2 Stimuli

Recordings of the word lists. Two male French native speakers (talker 1 and talker 2) without history of vocal pathology, were chosen for the recordings. The two speakers were from two different regions ("Pays de la Loire" west of France and "Lorraine," east of France). However, none of the participants in the validation experiment noticed or commented on the accent of the talkers.

The words were pronounced using a carrier sentence, as in the original MRT, with the target word embedded in the middle of the sentence, in order to prevent the effect of sentence-final intonation. The carrier sentence was "Le mot X doit être indiqué" ("The word X should be indicated").

The recordings were digitalized with a Teac LX10 at 48 kHz sampling frequency with a 16-bit resolution using a Brüel & Kjaer 1/2 in. microphone (type 4189) with a B&K preamplifier 2669 followed by B&K conditioner 5935 (Brüel & Kjaer Sound & Vibration Measurement, Naerum, Denmark). The recordings took place in an audiometric booth (IAC acoustics). The microphone was placed at a distance of 60 cm from the talker's mouth. All 300 sentences (50 lists of 6 words) were written on a computer screen placed in front of the talker and appeared one after the other every 3 s. It took an average of 2 s to pronounce a sentence. The experimenter visually checked through a window in the audiometric booth that the talker kept focusing on his task, and kept a constant distance between him and the microphone. Moreover, the signal waveform of the recordings was monitored in real time by the experimenter on a screen outside the audiometric booth.

Table 1. Lists composing the Fr-MRT.

	Word #1	Word #2	Word #3	Word #4	Word #5	Word #6
01	Bouche	Bouge	Boude	Boule	Bourre	Bouffe
02	Bête	Bêche	Baisse	Bègue	Belle	Benne
03	Gagne	Gare	Gaffe	Gage	Gale	Gamme
04	Vigne	Vide	Ville	Vice	Vive	Vise
05	Phare	Fâche	Femme	Fane	Face	Phase
06	Fiche	Fige	Figue	File	Fine	Fils
07	Dort	Dogue	Dock	Donne	Dope	Dote
08	Dire	Digue	Digne	Dîne	Dix	Dite
09	Pire	Pige	Pique	Pile	Pipe	Pisse
10	Peigne	Père	Pêche	Pèle	Peine	Pèse
11	Coque	Colle	Cogne	Cosse	Cote	Corps
12	Jure	Juche	Juge	Jules	Jupe	Jute
13	Cours	Coule	Coude	Couve	Coupe	Couse
14	Tanche	Tende	Tangue	Tank	Tempe	Tente
15	Teigne	Terre	Thèse	Telle	Thème	Teck
16	Tort	Toge	Toque	Tome	Tonne	Top
17	Mère	Messe	Mêle	Même	Mène	Mette
18	Sourd	Souche	Soude	Soûle	Soupe	Soute
19	Sage	Sache	Sac	Sale	Sape	Sas
20	Lange	Lande	Langue	Lampe	Lance	Lente
21	Lave	Lâche	Lasse	Laque	Lame	Latte
22	Ruche	Rude	Rhume	Rune	Russe	Ruse
23	Rogne	Roche	Robe	Rock	Rome	Rote
24	Chaude	Chauffe	Chaume	Chausse	Chauve	Chose
25	Manche	Mange	Mande	Mangue	Manque	Menthe
26	Ronde	Fonde	Monde	Ponde	Sonde	Tonde
27	Beurre	Cœur	Leurre	Meurt	Peur	Sœur
28	Rôle	Gnôle	Khôl	Môle	Pôle	Tôle
29	Feigne	Daigne	Baigne	Peigne	Saigne	Teigne
30	Chère	Gère	Fer	Guère	Mer	Père
31	Jette	Bête	Dette	Fête	Cette	Tête
32	Saine	Chêne	Gêne	laine	Rêne	Naine
33	Dame	Femme	Gamme	Rame	Lame	Pâme
34	Jappe	Cape	Zappe	Nappe	Pape	Tape
35	Rage	Gage	Cage	Nage	Page	Sage
36	Bave	Gave	Cave	Lave	Rave	Pave
37	Chiche	Riche	Fiche	Biche	Miche	Niche
38	Bise	Dise	Guise	Lise	Mise	Vise
39	Bile	Cil	Fil	Mille	Pile	Ville
40	Douche	Couche	Louche	Mouche	Souche	Touche
41	Rousse	Douce	Gousse	Mousse	Pouce	Tousse
42	Choque	Roque	Coque	Loque	Moque	toque
43	Comme	Gomme	Nomme	Pomme	Somme	Tomme
44	Rente	Sente	Vente	Fente	Lente	Pente
45	Rance	Chance	Danse	lance	Panse	Tance
46	Jure	Cure	Dure	Mure	Pure	Sûre
47	Bois	Quoi	doigt	moi	noix	loi
48	Crin	Grain	Frein	Train	Brin	drain
49	douer	nouer	louer	Vouer	Rouer	Jouer
50	Bail	Caille	Faille	Maille	Paille	Taille

Presentation of the stimuli. In the validation experiment, each target word was presented in noise, at a different SNR.

Six different SNRs were used, ranging from −16 to +4 dB in 4 dB steps (−16, −12, −8, −4, 0, and +4 dB). The masking noise signal was similar to the noise signal used in the original MRT (House *et al.*, 1965): uniform spectrum up to 500 Hz followed by a decay of 9 dB per octave. The noise spectrum was close to those of military environments (aircraft) (Gee, 2005). This was consistent with the objective to use the Fr-MRT for military radio communication systems. The target
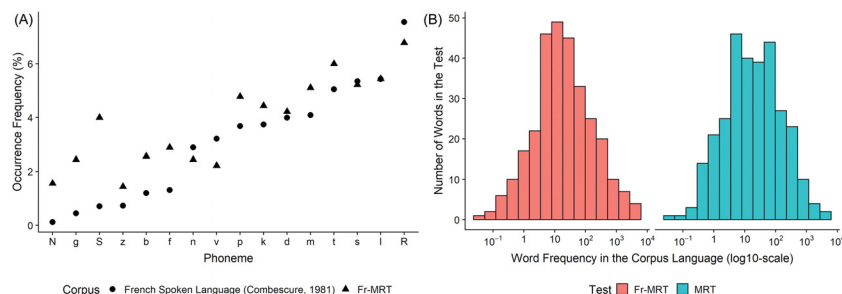
Fig. 1. (Color online) (A) Occurrence of consonants in French spoken language (Combescure, 1981) (●) compared to the occurrence of consonants in the Fr-MRT (▲). (B) Distribution of the 300 words used in Fr-MRT and MRT according to their frequency per million of occurrence in the corresponding spoken language. The log-scale was chosen to increase the readability of the figure.

word was emitted at 70 dB sound pressure level (Brüel and Kjær type 4157 ear simulator was used for calibration). The target word and the noise signal were presented simultaneously in the BeyerDynamic DT 770 headphone.

3.3 Procedure

Listeners sat in front of a computer in an audiometric booth. A trial began with the presentation of six written words on the screen. After the six words were displayed, the sentence with the target word was played on the headphone. Then, the listener had to indicate the word he/she heard, without time limit. The experiment was divided into two identical parts, corresponding to each talker (talker 1, talker2), with a 5-min break in between. The order of the two talkers was counterbalanced between the participants. For each talker, the listener performed six blocks of 50 trials, each trial corresponding to one list in the Fr-MRT. The six blocks corresponded to the six SNRs; they were presented in a random order.

Before data collection, the listeners performed a short training of ten-trial sets at 0 dB SNR. The experiment lasted approximately 60 min.

Finally, in order to assess a potential learning effect, a subgroup of 16 listeners repeated exactly the same experiment 30 days after.

3.4 Comparison between behavioral data and objective measurements

For each SNR, an acoustical measurement of speech intelligibility was computed and compared to the MRT results. This acoustical measurement was assessed by the STI (Houtgast and Steenken, 1971; Houtgast and Steeneken, 1982). STI is a numeric measure of speech intelligibility, varying from 0 (bad) to 1 (excellent). The STI of the headphone device used for the Fr-MRT listening tests has been measured for the six SNRs. The noise was the same as the one used to estimate the performance of the Fr-MRT list. A microphone was placed in the ear canal entrance of an artificial head and connected to the STI meter (AL1 acoustilyser from NTI). The STI uses a reference signal provided by NTI to simulate the modulations of speech in seven octaves. The artificial head was placed in the audiometric booth. The noise and the reference signal that was provided by the STI meter were played simultaneously in the headphone placed on the artificial head using a 4157 B&K ear coupler.

The Fr-MRT scores and the STI obtained for each SNR were compared to the data of a previous study (Brammer et al., 2011). In that study, the relationship between STI and MRT was presented as a curve. We first extracted from this curve the 6 MRT scores corresponding to the 6 STI values obtained for each of the SNRs of our study. Then, we compared the original MRT scores to the Fr-MRT scores measured during the first session (Day 1) corresponding to the same STI.

3.5 Statistical analysis

The effect of talker, SNR, and site of data collection was assessed using a repeated measures analysis of variance (ANOVA) on data measured during the first session (Day 1) with talker (talker1 vs talker2) and SNR (from −16 dB to +4 dB by 4 dB steps) as within factors, and site of data collection as between factor. Before conducting the ANOVA statistical analysis, the Fr-MRT scores were transformed from percent correct to rationalized arcsine unit to control for ceiling effect (Sherbecoe and Studebaker, 2004).

The learning effect was assessed using a repeated ANOVA in the subset of participants who did the two sessions, with sessions (Day 1 vs Day 30), talker, and SNR as factors. In the case of violation of sphericity, a Greenhouse-Geisser correction was applied.
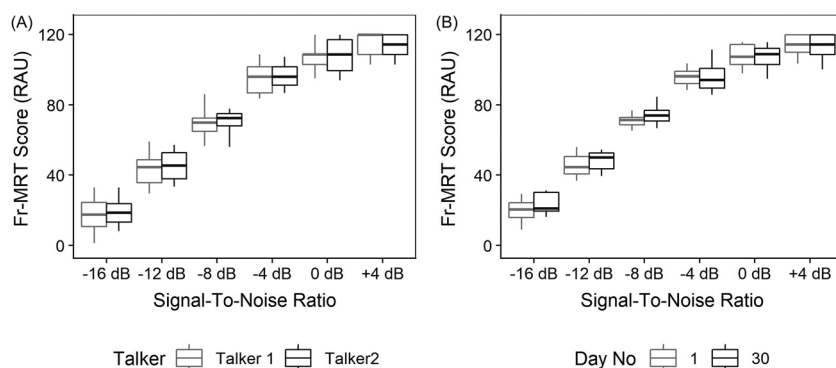
Fig. 2. (A) Fr-MRT scores (RAU, Rationalized Arcsine Units) of the listeners averaged across sites, plotted against SNR for Talker 1 (plotted in gray, left boxplots) and Talker 2 (plotted in black, right boxplots). The boxplots show the median (horizontal bar) and the interquartile range (boxes). The whiskers show the 5–95 percentile interval. (B) Fr-MRT scores (RAUs) of the listeners averaged across sites and talkers, plotted against SNR at Day 1 (plotted in gray, left boxplots) and Day 30 (plotted in black, right boxplots). The boxplots show the median (horizontal bar) and the interquartile range (boxes). The whiskers show the 5–95 percentile interval.

4. Results

4.1 Speech intelligibility as a function of SNR and talkers

As expected, the Fr-MRT scores measured at Day 1 increased monotonically with increasing SNR [$F(4.1, 97.2) = 1101$, $p < 0.0001$]. The difference of Fr-MRT scores between the two talkers was not significant [$F(1, 24) = 0.55$, $p = 0.46$], nor the interaction between the talkers and the SNR [$F(4.1, 97.2) = 0.20$, $p = 0.9$]. No effect of the site of data collection was observed [$F(1, 24) = 1.49$, $p = 0.23$], nor interaction between the site and the talkers [$F(1, 24) = 1.83$, $p = 0.18$], nor interaction between the site and the SNR [$F(4.1, 97.2) = 0.71$, $p = 0.89$], nor triple interaction [$F(4.1, 97.2) = 0.39$, $p = 0.82$]. Because of the absence of talker and site effects, the Fr-MRT scores were averaged across talkers and sites and used as a dependent variable in the next analyses. Figure 2(A) presents the collapsed results over site.

4.2 Learning effect

As for the MRT, there was no learning effect for the Fr-MRT: scores did not significantly change between Day 1 and Day 30 [see Fig. 2(B)].

4.3 Performance of the lists with initial consonant vs lists with final consonants

Finally, we analyzed the performance as a function of the place of consonant (initial vs final), in order to investigate potential differences between the different lists of the Fr-MRT. Figure 3 shows the performance averaged across listeners and talkers during the first session (Day1) for each list, according to the place of the consonant (initial vs final). The repeated measures ANOVA on the list performance with SNR as a within factor and the place of consonant as a
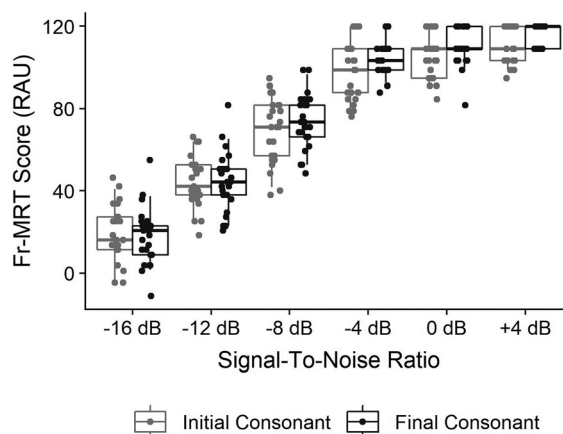

Fig. 3. Fr-MRT scores (RAUs) of the lists plotted against SNR for the lists with an initial consonant (plotted in gray, left boxplots) and the lists with a final consonant (plotted in black, right boxplots). The points indicate the average performance across listeners obtained for a list at a given SNR. The boxplots show the median (horizontal bar) and the interquartile range (boxes). The whiskers show the 5–95 percentile interval.
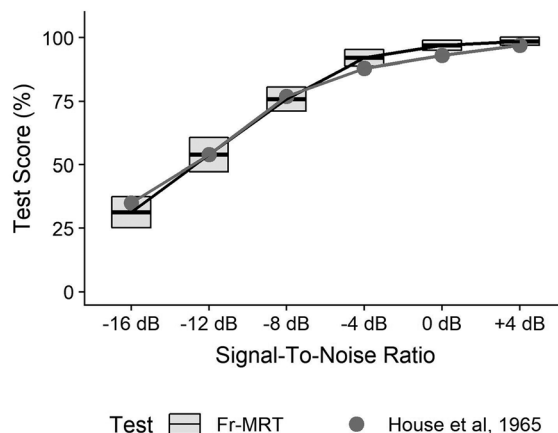
Fig. 4. Test scores (percent correct) averaged across listeners plotted against SNR for the Fr-MRT (crossbar) and the original MRT (gray points) (House *et al.*, 1965). The crossbars show the mean (horizontal bar) and the standard deviation (boxes).

between factor evidenced no effect of the place of the consonant [$F(1, 48) = 2.85$; $p = 0.09$], nor interaction between SNR and the place of the consonant [$F(4.1, 194.4) = 1.67$; $p = 0.14$].

4.4 Comparison with the original MRT study

The Fr-MRT and the MRT scores showed similar behavior across the SNR (see Fig. 4). Moreover, the average MRT score fell within the interquartile interval range of Fr-MRT scores for the three most adverse SNRs (-16, -12, -8 dB). For the other SNRs, the average Fr-MRT and MRT scores (Table III, part A; House *et al.*, 1965) were also close.

4.5 Relationship with the STI

Table 2 shows the STI and Fr-MRT scores obtained at each SNR, as well as the MRT score extracted from Brammer *et al.* (2011) corresponding to each STI value. The MRT scores fell close to the Fr-MRT scores for each condition.

5. Discussion

The goal of the present study was to design a French version of the MRT (here called Fr-MRT) to assess radio communication systems with French speaking users. As for the original MRT (House *et al.*, 1965), the Fr-MRT evidenced no learning effect (same performance 30 days after the first testing). Moreover, the variation of the Fr-MRT as a function of the SNR was very similar to the MRT one. Finally, the relationship between the Fr-MRT and the STI was also very similar to the one between the MRT and the STI (Brammer *et al.*, 2011). This is an important point because STI measurements are not possible in unconventional microphones for voice recording, such as bone conduction microphones or microphones integrated in earplugs. Hence, the development of the Fr-MRT enables to estimate the intelligibility for those specific microphone devices, with French speakers. The same reasoning could be applied in the future in the development of the test in other languages.

The previous adaptations of the MRT in a non-English language differed in their approach compared to the current study and, as a result, there is a lack of data to confront our results with. The Czech (Tihelka and Matoušek, 2004) and the Catalan (Alías and Triviño, 2007) MRTs did not provide any behavioral data. The Turkish MRT showed data of hearing-impaired listeners only, without estimation of neither a learning effect nor an effect of the SNR (Arıöz and Günel, 2015). The

Table 2. STI obtained for each SNR used in the current study, and the corresponding Fr-MRT scores (current study) and MRT scores (Brammer *et al.*, 2011) of each STI value.

SNR	STI	Fr- MRT Scores mean (sd)	Extracted MRT scores from Brammer *et al.* (2011)
−16 dB	0.05	31.3 (6.2)	30.9
−12 dB	0.1	54.0 (6.7)	50.2
−8 dB	0.25	75.9 (4.8)	76.2
−4 dB	0.38	92.1 (3.3)	87.6
0 dB	0.50	97.0 (2.0)	93.3
+4 dB	0.62	98.6 (1.58)	95.6

Spanish MRT assessed scores in normal hearing listeners but without an estimation of neither a learning effect nor an effect of the SNR (Ball, 2011). Here, we reproduced the experimental conditions of the original MRT (House *et al.*, 1965) in order to compare our results with the original ones. In addition, in French, most of the phonemes are shared with the English language. That characteristic has probably helped to build a Fr-MRT with similar properties as the original MRT.

Furthermore, the absence of a learning effect has interesting practical consequences: it allows native listeners to be tested with minimal familiarization, as underlined with the original MRT (House *et al.*, 1965).

Future work will employ female talkers to account for the increased number of military servicewomen.

Finally, we can easily imagine that French–English bilingual listeners could help to characterize the effects of a radio communication system on human voice features by comparing the MRT and Fr-MRT scores.

References and links

Alías, F., and Triviño, M. (**2007**). "A phonetically balanced modified rhyme test for evaluating Catalan speech intelligibility," in *Proceedings of 16th International Congress of Phonetic Sciences (ICPhS 2007)*, Saarbrücken, Germany, pp. 2197–2200.

Anderson, B. W., and Kalb, J. T. (**1987**). "English verification of the STI method for estimating speech intelligibility of a communications channel," J. Acoust. Soc. Am. **81**(6), 1982–1985.

ANSI (**1989**). ANSI S3.2-1989, *Method for Measuring the Intelligibility of Speech over Communication Systems* (American National Standards Institute, New York).

Arıöz, U., and Günel, B. (**2015**). "Modified Rhyme Test for evaluating Turkish speech intelligibility," Gazi Med. J. **26**(3), 75–81.

Ball, R. I. (**2011**). "A Spanish-language modified rhyme test for word recognition testing," Towson University Institutional Repository.

Barker, D. J., and Coyne, K. M. (**2017**). "Speech intelligibility during respirator wear," in *Handbook of Respiratory Protection* (CRC Press, Boca Raton, FL), pp. 93–112.

Bizzocchi, A. L. (**2017**). "How many phonemes does the English language have?," Int. J. Studies English Lang. Lit. **5**(10), 36–46.

Brammer, A. J., Yu, G., Bernstein, E. R., Cherniack, M. G., Tufts, J. B., and Peterson, D. R. (**2011**). "Intelligibility of speech corrupted by nonlinear distortion," in *Proceedings of the 10th International Congress on Noise as a Public Health Problem* (July), Vol. 24, pp. 287–292.

Brammer, A. J., Yu, G., Grady, J. J., Parham, K., Chernialk, M. G., Wannagot, S., and Cienkowski, K. M. (**2017**). "Are reports of temporary threshold shift-like symptoms by humans with normal hearing associated with hidden hearing loss?," J. Acoust. Soc. Am. **141**(5), 3814–3814.

Brandewie, E., and Zahorik, P. (**2011**). "Adaptation to room acoustics using the Modified Rhyme Test," Proc. Meet. Acoust. **129**(4), 2487.

Calliope, L. A., and Fant, G. (**1989**). La parole et son traitement automatique (Masson Paris). pp. 1–10

Combescure, P. (**1981**). "20 listes de dix phrases phonétiquement équilibrées," Rev. D'acoustique **56**, 34–38.

Gee, K. L., and Sparrow V. W. (**2005**). "Quantifying nonlinearity in the propagation of noise from military jet aircraft" ("20 lists of ten phonetically-balanced sentences"), in *Proceedings of the NOISE-CON*, Minneapolis, Minnesota (October 17–19).

Griffin, L. A. (**1992**). "Comparison of the speech transmission index and the modified rhyme test in simulated cockpit ambient noise," J. Acoust. Soc. Am. **91**(4), 2328–2328.

House, A. S., Williams, C. E., Heker, M. H., and Kryter, K. D. (**1965**). "Articulation-testing methods: Consonantal differentiation with a closed-response set," J. Acoust. Soc. Am. **37**, 158–166.

Houtgast, T., and Steeneken, H. J. M. (**1971**). "Evaluation of speech transmission channels by using artificial signals," Acustica **25**, 355–367.

Houtgast, T., and Steeneken. H. J. M. (**1982**). "A multi-language evaluation of the RASTI-method for estimating speech intelligibility in auditoria," Institue for Perception RVO-TNO.

Letowski, T. R., and Scharine, A. A. (**2017**). "Correlational analysis of speech intelligibility tests and metrics for speech transmission," U.S. Army Research Laboratory, Aberdeen Proving Ground, United States.

Manning, C., Mermagen, T., and Scharine, A. A. (**2017**). "The effect of sensorineural hearing loss and tinnitus on speech recognition over air and bone conduction military communications headsets," Hear. Res. **349**, 67–75.

MIL-STD-1474E. (**2015**). U.S. Department of Defense design criteria standard, noise limits, AMSC 9542, Washington, DC.

New, B., Pallier, C., Brysbaert, M., and Ferrand, L. (**2004**). "Lexique 2: A new French lexical database," Behav. Res. Methods Instrum. Comput. **36**, 516–524.

OANC (**2015**). Open American National Corpus, Open Data for Language Research and Education. http://www.anc.org/data/oanc/.

Pellieux, L., Bouy, J. C., Blancard, C., and Guillaume, A. (**2005**). "Speech intelligibility with a bone vibrator," RTO Human Factors and Medicine Panel (HFM) Symposium held in Amersfoort, The Netherlands (April 11–13, 2005).

Sherbecoe, R. L., and Studebaker, G. A. (**2004**). "Supplementary formulas and tables for calculating and interconverting speech recognition scores in transformed arcsine units," Int. J. Audiol. **43**, 442–448.

Tihelka, D., and Matoušek, J. (**2004**). "The design of Czech language formal listening tests for the evaluation of TTS systems," in Proceedings of LREC, Vol. VI, pp. 1099–2002, Lisbon, Portugal.

Yu, G., Brammer, A. J., Swan, K., Tufts, J. B., Cherniack, M. G., and Peterson, D. R. (**2010**). "Relationships between the modified rhyme test and objective metrics of speech intelligibility," J. Acoust. Soc. Am. **127**(3), 1903–1903.