

■ Development of the Test de Phrases dans le Bruit (TPB)

■ Élaboration du Test de phrases dans le bruit (TPB)

Josée Lagacé
Benoît Jutras
Christian Giguère
Jean-Pierre Gagné

Abstract

The Test de Phrases dans le Bruit, which consists of five French lists of 40 recorded sentences and a speech babble, was developed for use in evaluating speech perception in noise. The development of the sentence material was based on an approach that had previously been employed for the Speech Perception In Noise test. The key word familiarity of the sentences was tested, as well as the intelligibility in noise. Measures were also performed to obtain equivalent difference of scores between the high and low predictability sentences across the lists. Based on the results obtained with a subset of adult participants, it is believed that the sentence list sets that evolved from this work have the characteristics to be useful for the exploration of the underlying auditory and/or language-based origins of speech perception problems in noise among the Canadian French population. However, the present findings should be interpreted with caution as only individuals with normal hearing function participated in the experiments. The results may not apply to individuals with speech perception problems in noise. Additional evaluations of the psychometric properties of the test must be performed before its clinical application. Nevertheless, these preliminary findings suggest that further development of the Test de Phrases dans le Bruit is warranted.

Abrégé

Le Test de phrases dans le bruit, qui est composé de cinq listes de quarante phrases enregistrées en français et d'un bruit de verbiage, a été conçu pour évaluer la perception de la parole dans le bruit. L'élaboration des phrases a été effectuée en suivant une approche similaire à celle du *Speech Perception in Noise* (SPIN). La familiarité du mot clé de chaque phrase a été vérifiée, ainsi que le degré d'intelligibilité dans le bruit. Le niveau de prévisibilité des phrases a été mesuré afin de s'assurer que la différence de performance entre les phrases hautement prévisibles et faiblement prévisibles soit équivalente entre les listes. D'après les résultats obtenus avec un sous-groupe de participants adultes, on croit que les listes de phrases mises au point avec cet essai pourront être utiles à la recherche sur l'origine des problèmes auditifs ou linguistiques sous-jacents de perception de la parole dans le bruit parmi la population canadienne-française. Cependant, les résultats actuels devraient être interprétés avec prudence, car seulement des personnes avec une acuité auditive normale ont participé aux expériences. Les résultats pourraient ne pas s'appliquer aux personnes souffrant de problèmes de perception de la parole dans le bruit. Des évaluations supplémentaires des propriétés psychométriques du test doivent être effectuées avant son application clinique. Néanmoins, ces résultats préliminaires suggèrent que la poursuite de l'élaboration du Test de phrases dans le bruit est justifiée.

Key words: speech in noise tests, language-based competencies, linguistic context

Josée Lagacé, PhD
École d'orthophonie et
d'audiologie,
Université de Montréal,
Centre de recherche du
CHU Sainte-Justine
Montreal, Quebec
Canada

Benoît Jutras, PhD
École d'orthophonie et
d'audiologie,
Université de Montréal,
Centre de recherche du
CHU Sainte-Justine
Montreal, Quebec
Canada

Christian Giguère, PhD
Programme d'audiologie et
d'orthophonie,
Université d'Ottawa
Ottawa, Ontario
Canada

Jean-Pierre Gagné, PhD
École d'orthophonie et
d'audiologie,
Université de Montréal
Centre de recherche de
l'Institut gériatrique
universitaire de Montréal
Montréal, Quebec
Canada

Many individuals report difficulty understanding speech in noise. For some of them, their speech perception problems in noise can be explained by their audiogram. For others, the underlying nature of their difficulties is not as obvious. In these cases, a better understanding of the listening problems may improve service delivery.

The *Speech Perception In Noise* (SPIN) test was originally developed to assess how well individuals with acquired peripheral hearing loss utilize contextual linguistic information to understand speech in noise (Elliott, 1995; Kalikow, Stevens & Elliott, 1977). The original test materials consist of ten tape-recorded lists of 50 sentences mixed with a twelve-talker speech babble. When the SPIN test is administered, the listener is asked to repeat the final word (key word) for each sentence. In each list, half of the sentences are *highly predictable* (HP) as they contain contextual linguistic information that facilitates the identification of the key word (e.g., *The candle flame melted the wax*). The other half of the list is composed of *low predictability* (LP) sentences (e.g., *Paul can't discuss the wax*), which contain little contextual linguistic information (Kalikow et al., 1977).

The SPIN test was developed on the premise that speech perception involves at least two types of processes: 1) the auditory processing of the signal and, 2) the language-based processing of that information (Kalikow et al., 1977). According to Kalikow et al. (1977), the recognition of the final word of the HP sentences can be accomplished through one or both of these operations, while the recognition of the LP sentences key word depends mainly on the auditory processing of the signal. The level of the babble noise at which the test is conducted can be varied while presenting the different lists of the SPIN sentences. This test manipulation is relevant for determining the extent to which responses for each type of sentences are affected by the signal-to-noise ratio (SNR; Kalikow et al., 1977).

Since the two types of sentences of the SPIN test only differ by the semantic and syntactic content, it is possible to determine the extent to which the listener benefits from the contextual linguistic information by analyzing the difference of the performance for the HP and LP sentences (Kalikow et al., 1977). Although the use of linguistic contextual cues is only one component of the top-down processing involved in the speech recognition process, it is at least possible to measure this listener's competency with the SPIN test, which is not the case with the other available speech in noise tests.

The original version of the SPIN test and the SPIN-R test (the revised version of the SPIN test by Bilger, Nuetzel, Rabinowitz & Rzeczkowski, 1984) have been used in many studies to explore the underlying origins of the speech perception problems in noise. For instance, it has been employed in studies conducted among populations of younger and older adults with normal hearing sensitivity thresholds (Dubno, Ahlstrom & Horwitz, 2000; Humes, Burk, Coughlin, Busey, & Strauser, 2007; Kalikow et al., 1977; Pichora-Fuller, 2008; Pichora-Fuller, Schneider

& Daneman, 1995), adults with permanent hearing impairment (Bilger et al., 1984; Schum & Matthews, 1992), as well as adults with learning difficulties (Elliott & Busse, 1987). According to the results obtained with the SPIN test, the speech understanding difficulties experienced by these populations were related to underlying auditory deficits. On the other hand, comparisons of the results obtained by native listeners (listeners who learned American English from birth) and non-native listeners (listeners who learned American English later in life) on the SPIN test have led to different outcomes. The results revealed that the levels of noise at which speech is intelligible are significantly higher for the native listeners compared to the non-native listeners (Bradlow & Alexander, 2007; Florentine, 1985; Mayo, Florentine & Buus, 1997). It was also observed that the benefit from linguistic context is significantly greater for the native listeners compared to the non-native listeners.

The SPIN test provides a method of delineating the relative contribution of the auditory and the language-based function involved in speech understanding in noise. At this point, there is no test available in Canadian French that is comparable to the SPIN test. A simple translation of the SPIN test sentences would not have been valid because of the differences in the linguistic structure and vocabulary between English and French. It was therefore necessary to develop a French adaptation of the SPIN test. A similar approach to the one used for the development of the original version of the SPIN test was taken to establish the *Test de Phrases dans le Bruit* (TPB). This paper describes the development of the TPB, which consists of five French lists of forty recorded-sentences and a speech babble.

The Development of the TPB

The approach used to develop the test lists of the TPB involved the measurement of the intelligibility of the key words in noise (Experiment 1), the evaluation of the difference between the scores obtained on the HP and the LP sentences (Experiment 2) and the verification of the performance on the TPB at various SNRs (Experiment 3). The series of experiments that lead to the development of the preliminary version of the TPB is described below.

Development of the Speech Materials

According to Kalikow et al. (1977), to simplify the task and to minimize the influence of linguistic and memory skills, the type of response to be required from the subject has to be a single word response. As for the SPIN test, it was determined that the response word for the TPB would be the last word of the sentence. This type of response is also convenient for the examiner as the scoring simply requires matching the response with the final word of the test sentence (Kalikow et al., 1977). In order to further control the linguistic content of the sentences, another restriction was that the key word had to be a monosyllabic word. Moreover, all the sentences were constrained to contain six to eight syllables.

As opposed to the SPIN test, which was developed for unilateral presentation of the sentences and the babble

noise, a bilateral presentation mode was selected for the TPB. This option was chosen based on the poor ecological validity of unilateral presentation when testing speech in noise (Besing, Koenke, Abouchacra, & Letowski, 1998; Jerger, Greenwald, Wamback, Seipel, & Moncrieff, 2000).

Because the familiarity of the words influence their intelligibility when they are presented in noise (Elliott et al., 1979; Epstein, Giolas & Owens, 1968; Kalikow et al., 1977), all the key words chosen for the test material were selected from the MANULEX database (Lété, Sprenger-Charolles, & Colé, 2004). The MANULEX is a web database listing word frequency values for 48,886 lexical entries encountered in 54 French books used in European French elementary schools (Lété et al., 2004). No such large database was available for words used in Canadian French children literature. Monosyllabic words with a frequency of use within the range of 7.7 to 935.4 per million words were selected from the MANULEX. The initial pool consisted of 200 key words. Within the constraints previously noted, a set of 200 HP sentences was developed (e.g., *Elle met la nappe sur la table*), as well as a set of 200 LP sentences (e.g.: *J'ai acheté une nouvelle table*).

The resulting corpus of 400 sentences was analyzed by two grade 3 teachers (i.e., teaching children of eight to nine years of age), who were speakers of Canadian French, to confirm the naturalness of the sentences. The teachers were also invited to provide suggestions to improve the naturalness of the sentence where needed. They were asked to take into account that the TPB was to be used with children and adults.

Following the revision of the sentence naturalness, nine female native Canadian French speakers aged from 9 to 11 years completed a paper-and-pencil test to confirm the predictability of the sentences. The 400 sentences were listed on answer sheets with the key word deleted. Participants were instructed to fill in the blank with a word that they thought would most likely occur at the end. For each of the HP sentences, if none of the participants had written the intended key word, the sentence was reworked to be more predictable. For each of the LP sentence, if one participant had written the intended key word, the sentence was reworked to be less predictable.

It was determined that the sentences should be recorded by a female speaker because of the predominance of female educators and caregivers in children's education (Fallon, Trehub & Schneider, 2000). A female speaker of Canadian French who had previously participated in similar recording sessions was chosen to produce the 400 revised sentences. The sentences were recorded in a quiet recording room at the University of Montreal, with a digital video camcorder (Canon GL2, Canon Canada, Mississauga, ON L5T1P7) to which an external lapel microphone (Audiotechnica Pro70, Tokyo, Japan) was connected. During the recording session, the camera was positioned at approximately 2.5 meters in front of the speaker. The microphone was hanging from the ceiling, positioned at approximately 0.5 meters in front of the speaker. The speaker was instructed

to articulate each sentence as naturally and as clearly as possible. The recordings were then organized into 400 individual sentence files using the iMovie 4 software (Apple Canada, Markham, ON L3R 5G2). To ensure a uniform level across the stimuli, the key words were edited with the Cool Edit Pro software (Cool Edit Pro version 2.1, Adobe Systems Canada, Toronto, ON M8X 2X3) to be within ± 2 dB of the root mean square average level (68.3 dB SPL) of the 400 key words.

Since the key words were selected from the European French MANULEX database, a test of word familiarity was conducted with a group of children who were speakers of Canadian French. This verification was necessary because of the cultural differences between European and Canadian French. Five lists of 40 key words were developed for the familiarity test. The key words were all taken from the recorded LP sentences audio files to ensure, as much as possible, a similar accentuation on each word. The five lists of key words were burned to individual audio compact discs (CD). Forty children (19 girls and 21 boys) ranging from 5.5 to 7.4 years of age (average of 6.5 years) participated in this study. A parent of each participant signed the consent form and completed a questionnaire. Each participant was tested individually in a quiet room where ambient noise level did not exceed the specifications for hearing screening in schools (ASHA, 1997). A hearing screening at the intensity level of 20 dB hearing level (HL) was performed with a portable audiometer (Maico MA 41, Maico GmbH, 10587 Berlin, Germany; Beltone AE2, Beltone, Glenview IL 60026) with TDH-39 headphones (Telephonics, Farmingdale, NY 11735) prior to the experiment. All participants had normal hearing sensitivity at 500, 1000, 2000 and 4000 Hz bilaterally. The exclusion criteria for this study were any history of language disorders, otological problems, attention disorders or general learning delays. Four lists were presented monaurally to each participant (two lists per ear) via a CD player (Panasonic RX-D27, Mississauga, ON, L4W 2T3) connected to the portable audiometer set at 60 dB HL and a pair of headphones. The listener was instructed to report each word that was presented and to guess if necessary. A total of 160 words out of the 200 were correctly identified by over 80% of the participants. This suggested that the majority of the selected words were familiar to Canadian French children of five to seven years old.

Following the familiarity testing of the key words, 60 words were removed from the corpus on the basis of different considerations: (a) words with a frequency of use score of less than 10 per millions words (according to the MANULEX database) yielding a recognition score of less than 50%, (b) homonymous words like *boue* and *bout* and (c) words with different pronunciation across Canadian French communities (e.g., *zoo*, *oeuf*, *down*). This eliminated 120 sentences from the pool of recorded sentences because each key word appeared once in a HP and once in a LP sentence.

The remaining 280 sentences were divided into seven lists of 40 sentences, ensuring that the familiarity value

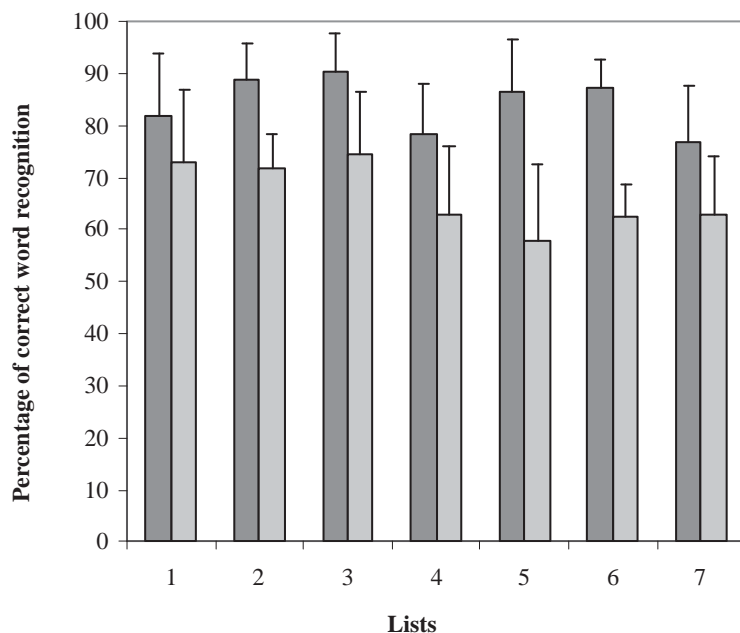


Figure 1: Group mean percent correct scores obtained by 10 adults for seven lists of sentences at a signal-to-noise ratio of 0 dB (Experiment 1). For each list, the dark grey bar represents the word correct score for the 20 high predictability sentences and the grey bar represents the word correct score for the 20 low predictability sentences.

of the key words was evenly distributed across the lists. Each list contained 20 HP and 20 LP sentences. A key word appeared only once in a given list, as in the SPIN test (Kalikow et al., 1977). The lists were transferred onto seven separate CDs for the speech intelligibility in noise testing described in the following section.

Experiment 1 - Measurement of the Key Words' Intelligibility in Noise

The goal of Experiment 1 was to determine if the speech intelligibility in noise of the key words was homogeneous across the seven sentence lists.

Participants

Ten Canadian French speaking adults (five females and five males) between 19 and 28 years of age (average of 22 years) were recruited for the measurement of the key words' intelligibility. Once the consent form was signed, each participant completed a questionnaire to rule out any exclusion criteria such as history of otological problems, language delay, attention disorders or general learning delay. If none of the exclusion criteria were identified, the participants were asked to undergo a bilateral hearing screening at 500, 1000, 2000 and 4000 Hz in an audiometric test suite. Using a Midimate 622 audiometer (GN Otometrics, Schaumburg, IL 60173 5329), the test tones were presented at 15 dB HL with TDH 39 headphones. If no sign of hearing loss was identified, the individual was invited to participate in the experiment.

Procedure

Each participant was tested individually in an audiometric suite, using the same audiometer and headphones as for the hearing screening. The sentences were transmitted via one CD player (Panasonic RX-D27) connected to the audiometer. The speech babble was conveyed via another CD player (TASCAM CD-A500, TEAC Canada, Mississauga, ON L4Z 1Z8) connected to a different audio-input channel of the audiometer. The seven lists of 40 sentences were presented at a SNR of 0 dB (the sentences and the speech babble at 65 dB HL) with monaural right ear presentation. The selection of the SNR of 0 dB was based on Kalikow et al.'s (1977) work for the SPIN test. The speech babble of European French talkers (4 females and 4 males) by Perrin and Grimault (2005) was used. Among the available pre-recorded babble, this was the most representative of the babble conditions of the target population (i.e., speakers of Canadian French). The speech babble was recorded in a continuous loop on a separate CD.

The order of the lists of sentences was partially counterbalanced across the participants (based on a Latin Square design). Participants were instructed to report the last word of each sentence they heard and to guess if necessary.

Results

The percent correct score average for the HP and LP items and standard deviations for each list are provided in Figure 1. Across the seven lists, the word recognition score ranged from 77% to 90.5% for the HP sentences (range of 13.5%), and from 58% to 74.5% for the LP sentences (range of 16.5%).

As for all the statistical analyses presented in this paper, an arcsine transform was applied to the data to stabilize the error of variance (Studebaker, 1985). An alpha level of 0.05 was used for all the statistical comparisons unless otherwise indicated. A repeated-measure, two-way analysis of variance (ANOVA) was performed on the mean average score obtained for the HP sentences and the LP sentences at each list. The ANOVA was conducted with the factor Type of sentences (HP and LP sentences) and the factor List (seven lists). There was a significant main effect of Type of sentence [$F_{(1,9)} = 98.73, p < .001, \eta^2 = 0.92$] across the seven lists. There was also a significant main effect of List [$F_{(3.49,31.38)} = 5.28, p < .001, \eta^2 = 0.37$]. The interaction of Type of sentences x List was significant [$F_{(6,54)} = 2.73, p = .022, \eta^2 = 0.23$]. This significant interaction was anticipated given that the HP-LP difference score ranged appreciably across the lists, e.g., from 9% to 28.5%. Because the sentence sets had to be re-worked to ensure an even distribution of the key words' intelligibility in noise values across the seven lists, no further statistical analyses were undertaken.

The sentences were re-assembled into a different set of

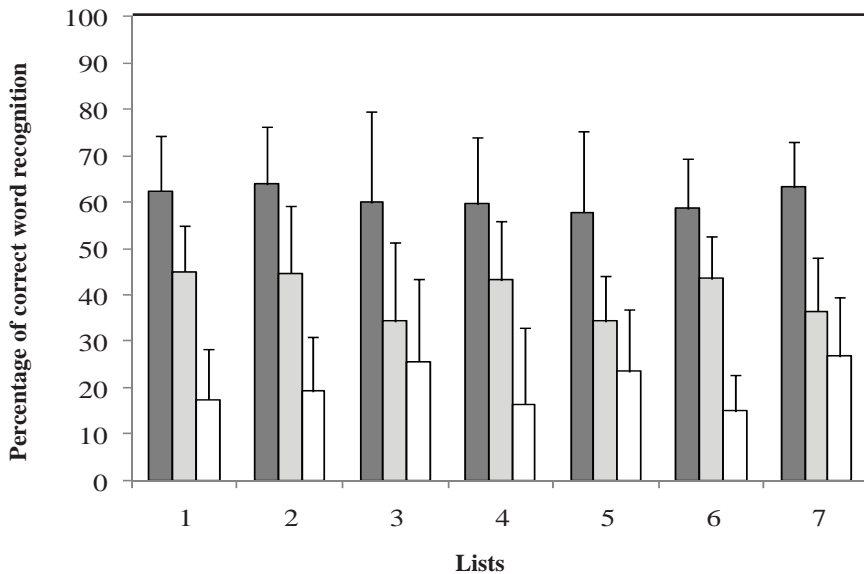


Figure 2: Group mean percent correct scores obtained by 14 adults for seven lists of sentences at an SNR of -2 dB (Experiment 2). For each list, the dark grey bar represents the word correct score for the 20 HP sentences, the grey bar represents the word correct score for 20 LP sentences and the white bar represents the mean of the difference scores between the HP and the LP sentences.

seven lists of 40 sentences, ensuring an even distribution of the key words' score across the lists according to their intelligibility values. Across the seven revised lists, the word recognition score ranged from 83% to 86.5% for the HP sentences (range of 3.5%), and from 67.5% to 65.5% for the LP sentences (range of 2%). To ensure that the revised lists were homogeneous, an ANOVA was performed on the mean average score of the HP sentences and the LP sentences of the revised lists. The results revealed no significant main effect of List [$F_{(6,114)} = .08, p = 0.998, \eta^2 = 0.00$]. The interaction of Type of sentences \times List was also not significant [$F_{(6,114)} = .02, p = 1.000, \eta^2 = 0.00$]. The revised lists were recorded on seven separate CDs for the evaluation of the difference of scores between the HP and the LP sentences across the lists, described in the following section.

Experiment 2 – Evaluation of the Difference of the Scores between the HP and the LP Sentences

As the aim of the TPB is to provide a mean for evaluating the extent to which the listeners can take advantage of the linguistic context, the lists had to be equivalent not only for the intelligibility in noise of the key words, but also for the difference of scores between the HP and the LP sentences. The goal of the Experiment 2 was to verify the equivalence of the seven revised lists and to ensure that the difference of scores between the HP and the LP sentences was homogenous across the lists.

Participants

A sample of 14 adults (11 females and 3 males) between 21 to 27 years of age (average of 23 years), speakers of Canadian French, was recruited for this study. None of the participants had taken part in Experiment 1. Prior to the experiment, participants were asked to sign the consent form and to complete a questionnaire to rule out any exclusion criteria. The inclusion and exclusion criteria used to recruit the participants were the same as Experiment 1.

Procedure

Each participant was tested individually in an audiometric suite with the same equipment as in Experiment 1. The seven revised lists of 40 sentences were presented at a SNR of -2 dB (sentences at 65 dB HL and speech babble at 67 dB HL) with monaural right ear presentation. The selection of the SNR of -2 dB was based on pilot data obtained from three participants. The pilot data indicated that the maximum difference in performance between the HP and the LP sentences was within that range of SNR. The same speech babble CD by Perrin and Grimault (2005), which was used in the Experiment 1, was employed for this experiment. The order in which the sentence lists were presented was partially counterbalanced across the participants (based on a Latin Square design). Participants were instructed to report the last word of each sentence they heard and to guess if necessary.

Results

The percent correct score average for the HP and LP items, as well as the difference of scores between the HP and the LP sentences, for each list are summarized in Figure 2. Across the lists, the scores ranged from 57.5% to 63.9% for the HP sentences and from 34.3% to 45% for the LP sentences. The average difference scores between the HP and the LP sentences ranged from 15% to 27%.

A repeated-measures two-way ANOVA was performed on the mean average score obtained for the HP sentences and the LP sentences for each list. The ANOVA was conducted with the factor Type of sentences (HP and LP sentences) and the factor List (seven lists). There was a significant main effect of the Type of sentences [$F_{(1,13)} = 11.72, p < .001, \eta^2 = 0.47$], but the main effect of List did not reach significance level [$F_{(6,78)} = 0.77, p = .60, \eta^2 = 0.06$]. The interaction of Type of sentences \times List was significant [$F_{(6,78)} = 29.3, p < 0.001, \eta^2 = 0.69$], indicating that the difference score between the two types of sentences was influenced by the list. The results of these analyses suggested that the lists were equivalent if considering the total average of correct recognition score (HP and LP sentences collapsed). However, when evaluating the average score obtained for the HP and the LP sentences separately, the

Table 1

Samples of the actual version of the TPB sentence lists. The type of sentences is indicated in parentheses at the end of each item, i.e., HP for the high predictable sentences and LP for the low predictable sentences. Each key word appears once in HP and once in a LP sentence, but only once in a given list, for example, the word “camp” appears in the list 1 in the HP context (bold) and in the list 2, in the LP context.

Liste 1

- | | |
|---|--|
| <ol style="list-style-type: none"> 1. Ce marchand vend des perles. (LP) 2. Claudie a découvert une mine. (LP) 3. Mon grand-père se berce sur sa chaise. (HP) 4. J'ai lu le livre jusqu'à la fin. (HP) 5. Il grave son nom sur du bronze. (LP) 6. Nos poumons respirent toujours de l'air. (HP) 7. Ma grand-mère a cousu ma robe. (LP) 8. J'ai quatre as dans mon jeu de cartes. (HP) 9. Jeanne se coupe les ongles. (LP) 10. Le chanteur a une très belle voix. (HP) 11. Tu as attaché ta tuque. (LP) 12. Elle lui fait signe de la main. (HP) 13. Certains soldats deviennent des fous. (LP) 14. Cette couverture est faite en laine. (LP) 15. J'enlève la neige avec une pelle. (HP) 16. Une main a quatre doigts et un pouce. (HP) 17. Mes enfants jouent avec une toile. (LP) 18. Ce cheval appartient au roi. (LP) 19. Ce joueur d'hockey fait des belles passes. (HP) 20. Le ballon roule vers le but. (LP) | <ol style="list-style-type: none"> 21. Ma cousine a trouvé un gros os. (LP) 22. Cette chanson s'appelle « vive le vent ». (HP) 23. J'ai acheté de la gomme. (LP) 24. On voit mieux la lune pendant la nuit. (HP) 25. Les deux garçons jouent à la guerre. (LP) 26. Il est mort quand j'avais cinq ans. (HP) 27. Ils chantent autour du feu de camp. (HP) 28. Les deux amis ont fait la paix. (LP) 29. Il faudra mettre une deuxième couche. (LP) 30. J'enferme mon chat dans sa cage. (HP) 31. Tous les trains roulent sur des rails. (HP) 32. Nous lui avons donné un verre. (LP) 33. Ce chandail n'a pas de prix. (LP) 34. Il fend le bois avec une hache. (HP) 35. Le frappeur a frappé la balle. (HP) 36. Maman a coupé les fleurs. (LP) 37. L'avion vole haut dans le ciel. (HP) 38. Le fermier va nourrir ces vaches. (LP) 39. Le jardinier arrose ses plantes. (HP) 40. Les girafes ont un grand cou. (HP) |
|---|--|

Liste 2

- | | |
|--|---|
| <ol style="list-style-type: none"> 1. Ce quilleur fait tomber toutes les quilles. (HP) 2. Jacinthe s'en va à son cours d'art. (LP) 3. Ils sont tous partis au camp. (LP) | <ol style="list-style-type: none"> 4. J'aime le beurre à l'ail. (HP) 5. ... |
|--|---|

statistical analyses revealed an effect of the list. This was an indication that the average of the difference scores between the HP and the LP sentences was not equivalent across the sentence sets. Repeated-measures one-way ANOVA was performed on the mean average of the difference scores (between the HP and the LP sentences) obtained for each list. The effect of list did not reach significance level [$F_{(6, 78)} = 1.98, p = .08, \eta^2 = 11.87$]. However, it was felt that a revision of the lists was necessary because of the range of the difference scores average across the lists, i.e., from 15% to 27%.

To obtain equivalent list sets and to maximize the difference scores between the HP and the LP sentences, individual analysis of each pair of sentences was performed. The percentage value of the recognition score obtained for the HP and LP sentences of each key word was compared. For some key words, the percentage value obtained for the HP and the LP sentence was similar. In other cases, the percentage value of the recognition score for the LP sentence was higher than the HP sentence (for the same

key word). In both instances, the pairs of sentences had to be eliminated from the corpus as the constraints were not met. A total of 80 sentences were removed from the corpus (i.e., 40 key words), based on these individual analyses.

The remaining 100 HP sentences and 100 LP sentences were assembled into five lists of 40 sentences, ensuring an even distribution of the key words according to their familiarity and intelligibility in noise values (from Experiment 1). Precautions were also taken for having equivalent means of the difference scores between the HP and LP sentences across the lists (from the results obtained in Experiment 2). The five lists were recorded on one CD, each list on a different track. This constituted the preliminary version of the sentence lists set of the TPB (see a sample of the TPB lists in Table 1). The performance at various SNRs had to be verified. This verification is described in the following section.

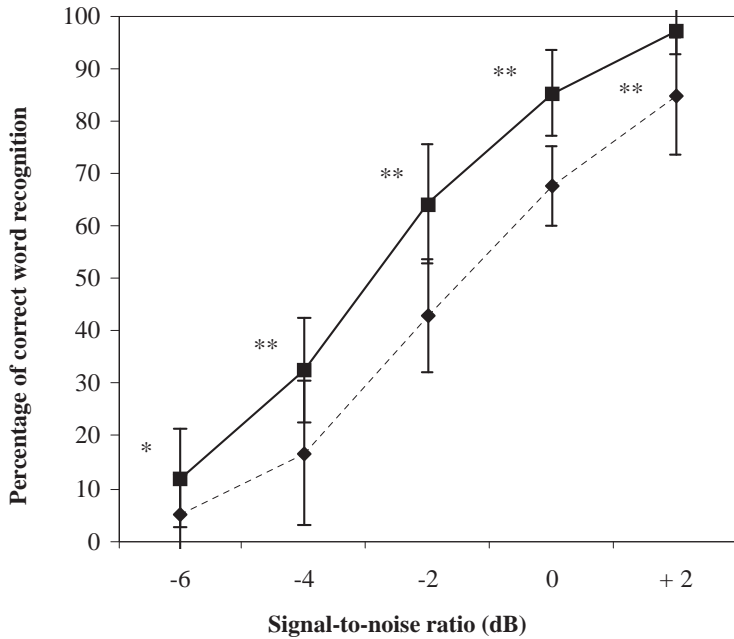
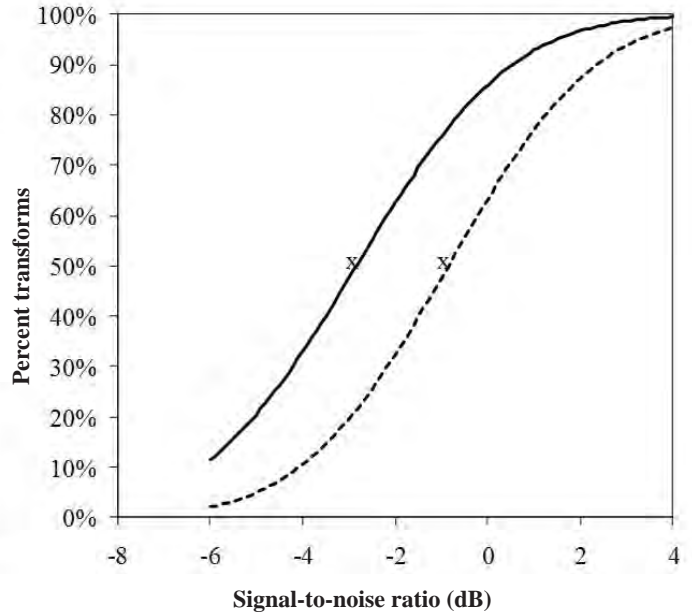
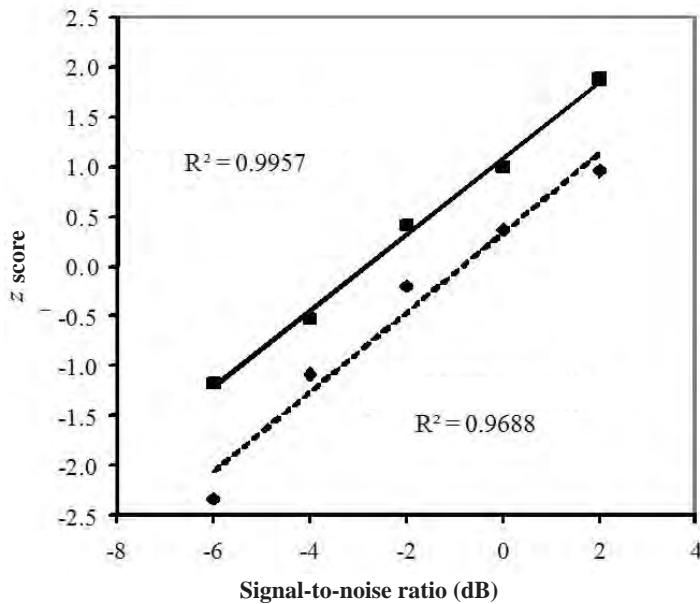


Figure 3 (left): Group mean percent correct scores obtained by 15 adults at various SNRs with the TPB sentence lists (Experiment 3). The solid line illustrates the performance with the HP sentences and the broken line illustrates the performance with the LP sentences at each SNR. Significant differences between both types of sentence are indexed with stars (* $p = 0.01$; ** $p < 0.001$).

Figure 4 (below):

The group mean percent correct scores illustrated in Figure 3 were transformed into z-scores. The z-scores as a function of the SNR for the HP sentences (square symbol) and the LP sentences (diamond symbol) are illustrated in the left panel. The linear regression function derived from the z-scores is illustrated with a solid line for the HP sentences and a broken line for the LP sentences. In the right panel, an Ogive (cumulative frequency) plot of target word intelligibility is shown, with a solid line for the HP sentences and a broken line for the LP sentences. The SNR at which a 50% key word intelligibility score would be reached is indicated by the symbol « x ».



Experiment 3 - Verification of the Performance on the TPB at Various SNRs

The objective of this experiment was to verify the performance on each type of sentence as a function of SNR.

Participants

A group of 22 Canadian French speaking adults was recruited for this study. None of them had taken part in Experiment 1 or 2. The age range extended from 19 to 43

years (average of 27 years). As for the previous experiments, all the participants were required to sign a consent form and to complete a questionnaire to rule out the presence of any exclusion criteria. The inclusion and exclusion criteria were the same as those used for the Experiment 1.

Seven participants had to be excluded from the study. Two participants reported a diagnosis of attention disorder during their childhood. One participant failed the audiological screening assessment. The data from four participants were discarded because they only completed

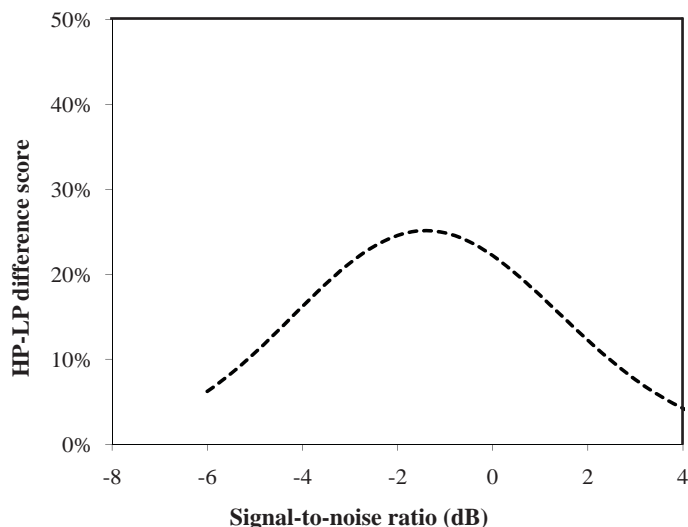


Figure 5: Difference of scores between the HP sentences and the LP sentences (in percent) as a function of signal-to-noise ratio (SNR), computed from the ogive functions illustrated in Figure 4.

four out of the five experimental conditions (due to lack of time). In total, the data from 15 participants (nine females and 6 males) were included in the analyses.

Procedure

This experiment was conducted in a quiet room at the Université du Québec à Trois-Rivières (Québec, Canada), which met the ANSI S3.1-1999 (R2008) specifications. The same equipment used for Experiment 1 and 2 was employed, i.e., Midimate 622 audiometer and TDH 39 headphones. The audiometer was connected to two Panasonic RX-D27 CD players. For this experiment, the testing conditions were similar to those of the TPB, i.e., bilateral presentation of the sentences and the noise. A local audiometric equipment company (Genie Audio Inc., Saint-Laurent, QC H4N 1T1) was consulted and developed an audio mixer that allowed bilateral presentation of both the test stimuli and the masking noise as well as an independent control of the intensity level of each one.

Each participant was tested individually with the five lists of the TPB and the same speech babble (Perrin & Grimault, 2005) as for the previous experiments. All the participants listened to each of the five lists presented at five different SNRs, i.e.: -6, -4, -2, 0 and +2 dB. The sentences lists were always presented at 60 dB HL. The order of presentation of the lists and SNRs was partially counterbalanced across the participants (based on a Latin Square design). The sentences and the babble noise were presented bilaterally.

Results

Mean results for the experiment are summarized in Figure 3. The percent correct word recognition scores for the HP and LP sentences obtained at each SNR are provided. The most consistent finding was that the mean average of the correct scores for the HP sentences was higher than

for the LP sentences percentage at the five tested SNRs. A repeated-measures two-way ANOVA was performed on the mean average score obtained for the HP and the LP sentences at each SNR to test the statistical significance of these trends. The analysis of variance was conducted with two within-subject factors: Type of sentences (HP and LP sentences) and SNR (five levels corresponding to the five tested SNRs). The analysis revealed a significant main effect of Type of sentences [$F_{(1,56)} = 268.35$, $p < .001$, $\eta^2 = 0.95$] and a significant main effect of SNR [$F_{(4,56)} = 273.97$, $p < .001$, $\eta^2 = 0.95$]. The interaction of the Type of sentence x SNR was also significant [$F_{(2,3,56)} = 8.46$, $p < .001$, $\eta^2 = 0.38$], suggesting that the difference score between the two types of sentences was influenced by the SNR. This was probably caused by the floor and ceiling effects of the performance-intensity function. For example, at -6 dB SNR, the performance for both LP and HP items approached 0%, reducing the difference of score between the two types of sentences.

Additional analyses were conducted to explore the nature of the Type of sentence x SNR interaction. This was accomplished by comparing the performance of the HP and the LP sentences at each SNR. Five paired *t*-tests indicated that the performance for the HP sentences were significantly different from the performance obtained with the LP sentences at each of the five SNRs tested, using the Bonferroni correction (critical alpha level of 0.01).

As in other studies on speech recognition performance (Boothroyd & Nittrouer, 1988; Laroche et al., 2003; Mayo et al., 1997), the mean percent correct scores for each type of sentences at each SNR was transformed into *z* scores. The *z* scores as a function of the SNR for each type of sentence is illustrated in the left panel of Figure 4. A linear regression function was calculated with the *z* scores. As shown in Figure 4, the data were well fitted by the linear regression function. The r^2 variance accounted for was over 0.9. The functions obtained for the HP and the LP sentences roughly showed a similar slope, i.e., 0.381 *z*/dB for the HP sentences and 0.400 *z*/dB for the LP sentences.

From the linear regression function, the *z*-scores were converted back to percentages to produce the intelligibility ogive (cumulative frequency) plots for the LP and HP sentences, as shown in the right panel of Figure 4. The data obtained with the HP and LP sentences of the TPB provided typical ogive speech intelligibility functions as the SNR increased. Based on these functions, it is noted that 50% key word intelligibility is reached at a lower SNR with the HP sentences (-2.8 dB) than with the LP sentences (-0.85 dB). This difference of SNR illustrates the contribution of the linguistic contextual information to auditory speech perception.

Difference scores were used in other studies to characterize the gain in speech perception performance attributable to the provision of additional linguistic and contextual cues (Elliott & Busse, 1987; Erber, 1975; Pichora-Fuller, 2008; Pichora-Fuller et al., 1995) or by the provision of visual information (Gagné, Tugby, & Michaud, 1991; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007). Therefore,

an HP-LP sentences difference score was computed from the ogive functions illustrated on the right panel of Figure 4. The plot of the HP-LP difference of scores as a function of SNR is illustrated in Figure 5. The plot reveals an inverted u-shaped relationship between the gain in recognition accuracy due to the HP sentences' additional linguistic contextual cues and the SNR. Using the difference score measure, it appears that the maximum benefit of the linguistic contextual cues for this group of listeners occurs at the center of the curve, at the SNR of -1.5 dB, with a gain of 25%. This observation is attributable to the particular test conditions used in the present experiment.

Discussion

This paper described the development of the TPB, which is a French adaptation of the SPIN test. SPIN-like tests provide a useful and time-tested way to measure speech recognition performance in the presence of background noise. The results obtained from the TPB may be analyzed from different perspectives, for example, by studying the difference between the scores on HP and LP sentences as a function of SNR. This perspective illustrates the contribution of language knowledge and ability to use the linguistic context of the HP sentences to understand speech (Elliott & Busse, 1987). It also shows at which SNR the listener benefits the most from the linguistic and contextual cues (Pichora-Fuller, 2008; Pichora-Fuller et al., 1995). For the group of adults who participated in this study, it appears that the maximum benefit of the linguistic contextual cues occurs at the SNR of -1.5 dB, with a gain of 25%. This observation is limited to the particular test conditions of the present study. However, in the case of listeners with hearing problems, the maximal difference score may fall at a different SNR because of the shift in the listener's performance-SNR curve and the possible difference in slope of the LP and HP curves. Moreover, in the case of listeners who cannot benefit from the linguistic contextual cues because of a language deficit, the magnitude of the difference scores between the HP and LP sentences may be lower.

The exploration of the speech perception problems experienced by individuals with auditory processing disorder (APD) counts among the applications of the TPB. The American Speech-Language-Hearing Association (2005) describes APD as difficulties in the perceptual processing of auditory information at the level of the central auditory nervous system. However, at the present time, the results of available studies have not specifically and unequivocally identified the underlying causes of the reported speech perception problems in noise reported by individuals with APD. If the underlying dysfunction in the case of APD is related to the auditory processing of the acoustic speech signal and not to the language-based processing, listeners with APD should be equally competent at using linguistic contextual cues at the TPB as individuals without listening problems. At present, many general intervention programs proposed for the rehabilitation and management of APD include procedures

to increase auditory closure abilities in order to improve the use of linguistic contextual information to facilitate speech perception in noise. Auditory closure refers to the recognition of complete words, or utterances, when only parts are spoken or heard (Delk, 1991). However, if the TPB results demonstrate that listeners with APD have similar auditory closure abilities as control groups, such intervention may not be required. Findings of this sort would guide the professionals involved with listeners presenting with APD to develop more effective intervention plans.

The TPB could also be used with other populations with speech perception problems in noise, to investigate their auditory closure skills which is not possible with other available French speech in noise tests. However, more testing with these clinical populations will be necessary to determine the diagnostic properties and accuracy of the TPB.

Additional evaluations of the psychometric properties and diagnostic usefulness of the TPB must be performed before it can be routinely applied in research and clinical applications. First, the equivalencies of the actual lists have to be measured in more detail. Second, in order to ensure that the TPB is appropriate for children, the test will have to be evaluated with that population. Normative data will also have to be collected for both the adult and children populations before its routine use, to allow comparison of performance measured with populations presenting with speech perception problems in noise. Moreover, for the data collection, additional validation for different dialects of French (other than Canadian French) may have to be undertaken, as the performance on the TPB may be influenced by the dialect, like any other speech perception test.

Conclusion

The objective of this paper was to describe the initial steps used to develop the TPB. The present findings should be interpreted with caution as only individuals with normal hearing function participated in the experiments. Additional evaluations of the psychometric properties of the test have to be performed before its clinical applications. Nevertheless, the preliminary findings suggest that further development of the TPB is warranted. The sentence lists that resulted from the research described here will be useful for the exploration of the underlying auditory and/or language-based origins of speech perception problems in noise for speakers of Canadian French. A better understanding of the perception of speech in noise may inform the development of more specific and effective intervention programs.

Acknowledgments

The authors wish to thank Andréa Bissonnette, Amélie Gaudreault, Mélanie Gagnon, Charlotte Ballet, Marie-Josée Levasseur, Gassia Jakmakjian, Yang Huang and Marie-Claude Chouinard for their assistance at various stages of the data collection. A special thanks to Anne-Marie Hurteau for agreeing to be the talker in our recordings,

as well as to all the participants at the various stage of the test development. The authors would also like to extend their thanks to the Conseil des Écoles Publiques de l'Est de l'Ontario and the Université du Québec à Trois-Rivières for their collaboration to this project.

Portions of this paper were presented at the 9th Congress of International Commission on Biological Effect of Noise in Mashantucket (July 2008) and at the Colloque international de réadaptation sur la surdité, la surdicécité et les troubles du langage et de l'audition in Montréal (June 2009).

This work was supported by a doctoral fellowship from the Centre de recherche du CHU Sainte-Justine and Fonds québécois de la recherche sur la nature et les technologies.

References

- American Speech-Language-Hearing Association. (1997). *Guidelines for Audiologic Screening* [Guidelines]. Retrieved on March 14, 2010 from www.asha.org/policy.
- American Speech-Language-Hearing Association. (2005). *(Central) Auditory Processing Disorders* [Technical Report]. Retrieved on March 14, 2010 from <http://www.asha.org/members/deskref-journals/deskref/default>.
- ANSI S3.1. 1999 (R2008). *Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms* ANSI S3.1. 1999. American National Standards Institute, New York.
- Besing, J., Koenke, J., Abouchakra, K., & Letowski, T. (1998). Contemporary approaches to audiological assessment in young children. *Topics in Language Disorders, 18*, 52-70.
- Bilger, R.C., Nuetzel, M.J., Rabinowitz, W.M. & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech, Language, and Hearing Research, 27*, 32-48.
- Bothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America, 84*, 101-114.
- Bradlow, A.R., & Alexander, J.A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America, 121*, 2239-2349.
- Delk, J.H. (1991). *Comprehensive dictionary of audiology* (4th edition). Maynards: The Laux Company.
- Dubno, J.R., Ahlstrom, J.B., & Horwitz, A.R. (2000). Use of context by young and aged adults with normal hearing. *Journal of the Acoustical Society of America, 107*, 538-546.
- Elliott, L.L. (1995). Verbal auditory closure and the speech perception in noise (SPIN) Test. *Journal of Speech, Language, and Hearing Research, 38*, 1363-1376.
- Elliott, L.L., & Busse, L.A. (1987). Auditory Processing by Learning Disabled Adults. In D. Johnson & J. Blalock (Eds.), *Adults with Learning Disabilities: Clinical studies* (pp.107-129). New York: Grune & Stratton.
- Elliott, L.L., Connors, S., Kille, E., Levin, S., Ball, K., & Katz, D. (1979). Children's understanding of monosyllabic nouns in quiet and in noise. *Journal of the Acoustical Society of America, 66*, 12-21.
- Epstein, A., Giolas, T.G., & Owens, E. (1968). Familiarity and Intelligibility of Monosyllabic Word Lists. *Journal of Speech, Language and Hearing Research, 11*, 435-438.
- Erber, N.P. (1975). Auditory-visual perception in speech. *Journal of Speech and Hearing Disorders, 40*, 481-492.
- Fallon, M., Trehub, S.E. & Schneider, B.A. (2000). Children's perception of speech in multitalker babble. *Journal of the Acoustical Society of America, 108*, 3023-3029.
- Florentine, M. (1985). Speech perception in noise by fluent, non-native listeners. *Journal of the Acoustical Society of America, 77*, S106-S106.
- Gagné, J.P., Tugby, K.G., & Michaud, J. (1991). Development of a Speechreading Test on the Utilization of Contextual Cues (STUCC): Preliminary Findings with Normal-Hearing Subjects. *Journal of the Academy of Rehabilitative Audiology, 24*, 157-170.
- Humes, L.E., Burk, M.H., Coughlin, M.P., Busey, T.A., & Strauser, L.E. (2007). Auditory Speech Recognition in Younger and Older Adults: Similarities and Differences Between Modalities and the Effects of Presentation Rate. *Journal of Speech, Language, and Hearing Research, 50*, 283-303.
- Jerger, J., Greenwald, R., Wamback, I., Seipel, A., Moncrieff, D. (2000). Toward a more ecologically valid measure of speech understanding in background noise. *Journal of the American Academy of Audiology, 11*, 273-282.
- Kalikow, D.N., Stevens, K.N., & Elliott, L.L. (1977). Development of a test of speech intelligibility in noise using materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*, 1337-1351.
- Laroche, C., Soli, S., Giguère, C., Lagacé, J., Vaillancourt, V., & Fortin, M. (2003). An Approach to the Development of Hearing Standards for Hearing-Critical Jobs. *Noise and Health, 6*, 17-37.
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX : A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, 36*, 156-166.
- Mayo, L.H., Florentine, M., & Buus, S. (1997). Age of Second-Language Acquisition and Perception of Speech in Noise. *Journal of Speech, Language and Hearing Research, 40*, 686-693.
- Perrin, F. & Grimault, N. (2005). *Fonds sonores*. Laboratoire Unités Mixtes de Recherche, Centre National de la Recherche Scientifique 5020, Lyon, France.
- Pichora-Fuller, K.M. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology, 47*, S72-S82.
- Pichora-Fuller, K.M., Schneider, B., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97*, 593-608.
- Ross, L.A., Saint-Amour, D., Leavitt, V.M., Javitt, D.C., & Foxe, J.J. (2007). Do You See What I Am Saying? Exploring Visual Enhancement of Speech Comprehension in Noisy Environments. *Cerebral Cortex, 17*, 1147-1153.
- Schum, D.J., & Matthews, L.J. (1992). SPIN test Performance of Elderly Hearing-Impaired Listeners. *Journal of the Acoustical Society of America, 3*, 303-307.
- Studebaker, G.A. (1985). A "Rationalized" Arcsine Transform. *Journal of Speech, Language and Hearing Research, 28*, 455-462.

Author's Note

Correspondence concerning this article should be addressed to Josée Lagacé, Audiology and Speech-Language Pathology Program, University of Ottawa, Roger Guindon Hall, 451 Smyth Road, Ottawa, Ontario, K1H 8M5. Email: Josee.Lagace@uottawa.ca.

Received: October 1, 2009

Accepted: March 24, 2010

