Speech audiometry in noise: Development of the French-language VRB (vocale rapide dans le bruit) test

F. Leclercq, C. Renard, C. Vincent

Laboratoire Dehaussy-Renard, rue nationale, 59000 Lille, France
Service d’otologie et d’otoneurologie, CHRU de Lille, 59000 Lille, France

ARTICLE INFO

Keywords:
Speech-in-noise perception
QuickSIN™
Speech-in-noise audiometry

ABSTRACT

Introduction: Understanding speech in noise is a major challenge for most hearing-impaired subjects, with or without hearing aids. To overcome the weaknesses of French-language speech-in-noise tests, we developed a new instrument, with a balanced mix of difficulty of the speech material.

Material and methods: The speech material comprised 127 sentences taken from the “Marginal Benefit from Acoustic Amplification” (MBAA) corpus, each including 3 keywords. The noise was created using the “onde vocale globale” (global vocal wave: OVG), described by Dodelé. The 127 speech/noise pairs were recorded individually after root-mean-square equalization. The first experiment, on 10 normal-hearing adults, determined the signal-to-noise ratio (SNR) associated with 50% correct keyword identification in each sentence (SNR-50), using an ascending method with noise level set at 73 dB SPL. Relative levels between sentences and noise were then adjusted sentence by sentence to achieve an SNR-50 of 0 dB. The second experiment, with 12 normal-hearing adults, validated the equalization of sentence difficulty.

Results: Mean SNR-50 was −6.64 dB (σ = 1.47). Mean adjusted SNR-50 was 0.08 dB (σ = 0.15). Mean psychometric curve slope was 19.3%/dB, with low standard deviations, testifying to the sensitivity of the speech material.

Conclusion: The VRB (vocale rapide dans le bruit: rapid speech in noise) test is based on sentences from the MBAA corpus with background noise based on the OVG at different signal-to-noise ratios. The test is feasible and able to detect slight variations in speech-in-noise performance between subjects.

© 2018 Elsevier Masson SAS. All rights reserved.

1. Introduction

As early as 1970, Carhart and Tillman argued that hearing difficulty should be assessed more ecologically than by pure-tone and speech audiometry in silence, which needs supplementing by testing speech perception in noise. Many authors, including among others Killion and Niquette (2000), Taylor (2003) and Wilson, McArdle and Smith (2007), reiterated this recommendation [1–3]. In France, however, speech-in-noise audiometry (SNA) is still not widely used.

Difficulty in understanding speech in noise is often one of patients’ main complaints. Good practice should systematically search for relevant information to objectify and quantify this complaint.

SNA has many advantages for hearing-aid adjustment, in selecting the device, assessing the usefulness of stereophonic aids, the efficacy of speech-processing algorithms and of directional systems and the contribution of associating an FM system [2]. In hearing-aid follow-up, SNA can also monitor performance resulting from hearing-aid fitting.

As there are few standardized French-language SNA tests, the present study aimed to develop a new one, the “vocale rapide dans le bruit” (rapid speech in noise: VRB), with homogeneous difficulty for normal-hearers, using a methodology similar to that of the American QuickSIN™ [4].

The SIN test [5] was at the origin of the QuickSIN. It was developed to assess speech intelligibility in noise with and without hearing aids. The speech material was taken from the IEEE (Institute of Electrical and Electronics Engineers) sentence list, each sentence having 5 keywords that are not highly predictable from context. The noise was of the multitalker type (3 female and 1 male voice). The lists comprise 40 sentences testing 4 SNRs, from +15 dB to 0 dB by 5 dB steps. In each list, 20 sentences are presented at a fixed level...
of 70 dB HL and the others at 40 dB HL. The drawbacks of the SIN are:

- long administration time (>10 min);
- 40 dB HL level too low for hearing-impaired subjects;
- lists not mutually equivalent.

QuickSIN™ [4,6] is an improved version, giving a quick estimate of the 50% correct-response SNR, on a descending method from least to most difficult. The lists comprise 6 IEEE sentences, of equivalent difficulty. Within each list, 6 SNRs are tested, from +25 dB to 0 dB by 5 dB steps. The Spearman-Kärber equation determines the 50% correct-response SNR (SNR-50), using a non-parametric method to calculate the 50% level of a psychometric function (here, intensity/SNA performance). It is an alternative to adaptive audiometric techniques.

The aim of the present study was thus to develop French-language speech material of homogeneous difficulty, to be used to develop an SNA test. The methodology consisted in:

- selecting a set of sentences and a noise meeting certain criteria;
- measuring the SNR-50 for each sentence;
- equalizing difficulty across sentences;
- validating difficulty equivalence for the speech/noise pairs;
- and creating three VRB test-lists and testing them in clinical practice.

2. Material and methods

2.1. Speech material and noise

The test sentences were taken from the Marginal Benefit from Acoustic Amplification (MBAA) corpus of 540 sentences, typical of everyday conversation; length varies greatly, from 3 to 15 words, and they are of all forms: affirmative, negative, interrogative and exclamatory. The speaker was a female speech-therapist, speaking French at normal speed without regional accent.

One-hundred-and-twenty-seven sentences were selected on the following criteria:

- short enough to be easily repeated [7]: only 7-11 syllables;
- not highly predictable from context, so as to limit mental compensation [4];
- not liable to create confusion: interrogative and exclamatory sentences in the MBAA corpus were not used [7];

- and containing at least three 1-, 2- or 3-syllable keywords for grading purposes; these could be of any type except prepositions, articles, pronouns, conjunctions or proper names [8].

The noise comprised an 8-second loop taken from the “onde vocale globale” (global vocal wave: OVG). This consists of 4-speaker noise composed from recordings of two couples, one French- and one English-speaking, in conversation [9]. The idea of pairing exactly the same noise with all sentences came from the WIN test [10,11], and was intended to reduce variability and facilitate synchronization between speech and noise in processing the data files.

2.2. Subjects

Sampling was performed on a predefined population. An upper age limit reduced age-related intelligibility bias, and a limit on age and educational level ensured adequate knowledge of French.

Participant inclusion criteria comprised:

- native French speaker;
- hearing thresholds < 20 dB HL at all octaves between 125 and 8000 Hz in both ears (Interacoustics AC-40 audiometer and TDH-39 headphones in soundproof booth);
- and normal otoscopy with no otologic history.

Experiment 1, determining SNR-50 for the sentences, included 10 normal-hearing French adults, 5 male and 5 female, aged 21–37 years (X= 27.7 years; σ = 4.95).

Experiment 2, confirming homogeneous sentence difficulty, included a new sample of 12 normal-hearing French adults, 8 male and 4 female, aged 19–39 years (X= 26.33 years; σ = 6.73).

VRB test feasibility in clinical practice was assessed in 29 hearing-impaired subjects aged 42–90 years (X= 67.72 years; σ = 15.02) and 6 normal hearing subjects aged 25–33 years (X= 30.17 years; σ = 3.97), using 3 lists of 9 sentences. No exclusion criteria were applied.

2.3. Protocol

Subjects were tested in a soundproof booth with TDH-39 headphones, with binaural diotic listening. Noise was set at 73 dB SPL, measured on a Brüel and Kjær 2270 sonometer-analyzer and Brüel and Kjær 4152 artificial ear. This level was chosen to represent noise levels generally encountered in social gatherings, which typically range between 65 and 85 dB SPL [4].
To familiarize subjects with the task while minimizing training effects, the test session began with a single training run of 10 sentences. Sentences were then delivered 10 by 10. In each sentence repetition, only the keywords were counted.

In experiment 1, list delivery began at $-14$ dB SNR; exploratory tests showed that no keywords were recognized below this level. After these first 10 sentences, delivery was repeated at $-12$ dB SNR, then increasing by 2 dB steps until all 3 keywords in the 10 sentences were correctly identified. List presentation ordered varied for each subject.

Task instructions were: “Imagine you are at a party. A woman speaks to you, but you are too far away to hear. You decide to get closer so as to understand better. As soon as you can hear her voice, repeat each word or sentence you hear. The background noise will make this difficult, but you should try to repeat or guess as many sentences or words as you can.”

The SNR-50 values obtained in experiment 1 served as reference to equalize difficulty across sentences. For example, for a speech/noise pair with SNR-50 of $-8.35$ dB, noise level was kept constant and speech level was reduced by $8.35$ dB so as to have a theoretic SNR-50 of 0 in experiment 2. SNR-50 was set at 0 dB for the normal-hearing group for purposes of thresholding, 0 dB corresponding to 50% correct responses on clinical audiometry; as SNR-50 is 0 for a normal-hearing subject, the SNR-50 obtained for a hearing-impaired subject will directly correspond to SNR loss.

For experiment 2, the protocol was the same, except that the SNR test range began at $-8$ dB. Like in experiment 1, a training run was used and sentence presentation order varied.

Test feasibility was assessed on 3 lists of 9 sentences. The VRB test used an automated descending procedure based on QuickSIN™. Each list tested 9 SNRs from +24 dB to 0 dB by 3 dB steps. Administration time for the 3 lists was 3 minutes 36 seconds, with 9 keywords per level: e.g., 81 items. Like in QuickSIN™, the VRB test expresses SNR-50 in dB using the Spearman-Kärber equation; the main difference is in the calibration of the VRB, setting normal-hearing SNR-50 at 0 dB, so as to assess an “SNR loss” with respect to normal values.

2.4. Scoring

Each correctly repeated keyword at a given SNR level scored 1 point. SNR-50 was calculated on the Spearman-Kärber equation.
This statistical method, first described by Spearman then by Kärber, calculates the 50% point for a psychometric function [12]:

$$T_{50\%} = i + \frac{d}{2} - \frac{d \times r}{n} = \text{SNR} - 50$$

where $T_{50\%}$ is the 50% point of the curve, $i$ is the initial test presentation level, $d$ the step difference in dB between levels, $r$ the number of correct responses, and $n$ the number of items tested per level.

### 3. Results

Before equalization of sentence difficulty, SNR-50 varied by 8.54 dB, with a mean of −6.64 dB ($\sigma=1.47$ dB). Fig. 1 shows SNR-50s per sentence, before and after equalization. Experiment 2 provided the new SNR-50s for each sentence, with variation down to 3.34 dB, for a mean $+0.08$ dB ($\sigma=0.35$ dB); e.g., approximating zero.

Fig. 2 shows psychometric curves before and after equalization. After equalization, the standard deviation for SNR-50 decreased from 1.47 dB to 0.55 dB.

Mean psychometric curve slope between 30% and 70% correct response levels was 19.3%/dB.

Fig. 3 shows exploratory test results for the 3 VRB lists in 29 subjects.

Mean SNR loss on the 3 VRB lists was 0.2 dB in the normal-hearing group, 4.6 dB in mild hearing loss, and 6.3 dB in first-degree moderate hearing loss on the BIAP classification and 10 dB in second-degree moderate hearing-loss, and 12.5 dB in severe hearing loss.

### 4. Discussion

The mean SNR-50 value of $-6.64$ dB in experiment 1 was close to those found in developing other sentence-based French-language tests: FrMatrix [13] and the FIST test [14] had mean SNR-50 values of respectively $-6.00$ dB and $-7.40$ dB. Comparison, however, is in fact difficult, due to differences in type of noise, methodology and scoring unit (sentence or keywords).

The present SNR-50 values varied considerably, over a range of 8.53 dB. This was close to the 7.20 dB range reported in developing the QuickSIN™. In both cases, this may be due to some sentences with their keywords being easier than others [4].

To equalize difficulty for the whole set of speech/noise pairs, SNR-50 should have a constant value. There are two means of achieving this: firstly, altering the total RMS value in one of the two audio file channels; or secondly, simply excluding outlying pairs with SNR-50 values differing greatly from the mean [8,15]. As a sufficient number of sentences were needed, we did not use the second method, as it would have excluded sentences that could be used with the first.

The value to which SNR-50 should be equalized has also been a matter of debate, with two approaches. In the literature, mean SNR-50 is often used. However, we decided to proceed differently: SNR-50 values were adjusted to 0 dB, so that the speech material could be used in a VRB test directly providing individual SNR loss with respect to normal values.

So as to have enough sentences to create the VRB lists, 120 were selected and equalized for difficulty; only 7 were excluded for atypical psychometric curves.

After adjusting speech/noise pair difficulty from experiment 1 to experiment 2, the range, mean and standard deviation of SNR-50 should be 0 dB; the present value of $+0.08$ dB was very close and not significantly different ($Z_{obs} = 1.59, P \leq 0.05$). It is also noteworthy that, between the two experiments, the range of SNR-50 values was reduced 2.55-fold and the standard deviation 2.67-fold.

Sentence psychometric curve slope is another important variable, directly reflecting the sensitivity of the speech material. A steep slope means that any slight change in RSB entails a wide variation in score [15]. Slopes of 15–20%/dB are often considered sufficient for an SNA test [8,16].

### 4.1. Experience 2

Mean slope was the same in the two experiments, at 19.3%/dB. In comparison, the literature reports mean slopes of 20.2%/dB for FIST [14] and 14.00%/dB for FrMatrix [13]; in English-language tests, slope was 11.9%/dB for BKB-SIN and 10.8%/dB for QuickSIN™ [3].

Administration of the 3 VRB lists was feasible, without difficulty in any subject. Results showed good sensitivity in distinguishing normal-hearing and hearing-impaired subjects: the former had SNR losses tightly clustered around 0 dB, the latter showing losses increasing with pure-tone threshold loss. However, the small number of subjects and lists calls for caution. Killion and Niquette [1] used data from 4 studies to demonstrate the inability of pure-tone audiometry to predict speech perception in noise; thus, patients with the same mean pure-tone threshold loss may show SNR-50 values varying by more than 10 dB.

### 5. Conclusion

The present study validated new speech material, balanced in difficulty, that can be used for creating a speech-in-noise audiometry test.

In line with the literature, various steps were needed for this study, resulting in 120 speech/noise pairs of equal difficulty, with SNR-50s clustering within a range of just 3.34 dB around a mean $+0.08$ dB ($\sigma=0.55$ dB).

These sentences are now available for creating SNA test lists. A preliminary version, the VRB, comprising 3 lists of 9 sentences, was created and seemed to validate the specificity and sensitivity of the speech material.

For the future, a larger number of lists will need drawing up. It should also be noted that using just THD-39 headphones or the equivalent is insufficient, and would not allow development of a test for the purposes of clinical diagnosis and hearing-aid...
assessment. It is therefore planned to adapt the test for free field conditions, enabling efficacy monitoring for patients with hearing aids or cochlear implants.

**Disclosure of interest**

The authors declare that they have no competing interest.

**Acknowledgements**

We thank Pr. M. Marx and Dr. C. James Purpan University Hospital, Toulouse, France, for permission to use the MBAA corpus, and Mr. Léon Dodelé for permission to use the OVG noise.

**References**