ISSN: 2083-389X
eISSN: 2084-3127

© J Hear Sci, 2022; 12(3): 9–21
DOI: 10.17430/JHS.2022.12.3.1

A WITHIN-SUBJECT COMPARISON OF HEARING AID PERFORMANCE IN NOISE BASED ON VERBAL RESPONSE TIMES

Horacio Cristiani[AC-G], **Sabrina Alonso**[ABDF]

Mutualidad Argentina de Hipoacúsicos, Buenos Aires, Argentina

Corresponding author: Horacio Cristiani, Mutualidad Argentina de Hipoacúsicos, Peron 1654, C1037ACF, Buenos Aires, Argentina; email: hcristiani@mah.org.ar

Contributions:
A Study design/planning
B Data collection/entry
C Data analysis/statistics
D Data interpretation
E Preparation of manuscript
F Literature analysis/search
G Funds collection

Abstract

Background: Verbal response times (VRTs) are among the suggested markers for cognitive load during word recognition tasks. Measurements of VRT during hearing aid fitting can be a useful tool to obtain information about listening effort with different amplification parameters.

Material and methods: A software program was developed to easily measure VRTs in speech recognition tests. The system plays 50 randomly chosen recorded words out of a set of 700 disyllables. Speech material can be presented together with pre-selected noise samples at different speech-to-noise ratios and processed with low-pass filters with selectable cut-off frequencies. The test is carried out in free field. A voice activity detector measures the time between the offset of the presented word and the onset of the repetition by the subject, which allows VRT and speech recognition scores to be quickly assessed. Tests were carried out with a group of 8 normal-hearing subjects to evaluate the effect of different filter parameters and a second group of 8 normal-hearing people to evaluate the effect of different speech-to-noise ratios on VRT. Finally, a group of 15 adult hearing-impaired subjects who used hearing aids were fitted under different conditions and the VRTs were compared between fittings.

Results: Reducing the low-pass filter cutoff frequency or adding noise to the speech signal increased VRTs in normal hearing people, suggesting an inverse relationship between VRT and ease of listening. In the hearing-impaired group, VRTs with different fittings of the hearing aid showed differences that can be used as an indicator of listening effort.

Conclusions: Adding a measurement of VRT to a regular word recognition test during hearing aid fitting could be useful for adjusting parameters or deciding between models or processing strategies, especially if recognition scores are high.

Keywords: speech recognition • listening effort • hearing aid fitting • verbal response time

WEWNĄTRZGRUPOWE PORÓWNANIE SKUTECZNOŚCI APARATÓW SŁUCHOWYCH W SZUMIE NA PODSTAWIE CZASÓW ODPOWIEDZI WERBALNEJ VRT

Streszczenie

Wprowadzenie: Czas odpowiedzi werbalnej (VRT) jest jednym z sugerowanych markerów obciążenia poznawczego podczas wykonywania zadania rozpoznawania mowy. Pomiar VRT w trakcie ustawiania aparatu słuchowego może być użytecznym narzędziem do uzyskania informacji o poziomie wysiłku słuchowego przy różnych parametrach wzmocnienia.

Materiał i metoda: Opracowano program komputerowy do ułatwienia pomiaru VRT w testach rozpoznawania mowy. Program odtwarza 50 słów dwusylabowych wybranych losowo ze zbioru 700 nagranych. Materiał słowny może być odtwarzany razem z wybranymi próbkami szumu, z różnym stosunkiem sygnału do szumu, z filtrem dolnoprzepustowym umożliwiającym wybór częstotliwości odcinającej. Test jest wykonywany w wolnym polu. Detektor aktywności głosowej mierzy czas pomiędzy początkiem prezentowanego słowa a początkiem powtarzania tego słowa przez osobę badaną, co umożliwia szybką ocenę VRT i poziomu rozpoznawania mowy. Grupa 8 normalnie słyszących osób wzięła udział w badaniu wpływu różnych parametrów filtrowania. W innej ośmioosobowej grupie normalnie słyszących osób oceniano wpływ zmiany poziomu stosunku sygnału do szumu na VRT. Na koniec w grupie 15 dorosłych osób z niedosłuchem korzystających z aparatów słuchowych wykonano ustawienie parametrów stymulacji w różnych warunkach i porównano VRT przy różnych ustawieniach.

Wyniki: Zmniejszenie częstotliwości odcinającej filtra dolnoprzepustowego lub dodanie szumu do sygnału mowy zwiększało VRT u osób z normalnym słuchem, co sugeruje istnienie odwrotnej zależności pomiędzy VRT a łatwością słyszenia. W grupie z niedosłuchem zaobserwowano zmiany VRT przy różnych ustawieniach aparatu słuchowego, które można uznać za wskaźnik poziomu wysiłku słuchowego.

Wnioski: Uzupełnienie powszechnie stosowanego testu rozpoznawania mowy o badanie VRT podczas dopasowania aparatu słuchowego może być użyteczne do lepszego dobrania parametrów stymulacji lub wyboru modeli lub strategii przetwarzania, szczególnie w przypadkach, gdy poziom rozpoznawania mowy jest wysoki.

Słowa kluczowe: rozpoznawanie mowy • wysiłek słuchowy • programowanie aparatu słuchowego • czas odpowiedzi werbalnej

Introduction

The concept of listening effort is receiving increasing interest in the field of hearing aid fitting. The cognitive resources consumed in word recognition tasks cannot be revealed by conventional speech audiometry. Understanding acoustically degraded speech, either in difficult environmental conditions or through hearing loss, requires additional cognitive assistance [1]. Greater acoustic challenges are associated with higher error rates in understanding speech, poorer performance on concurrent secondary tasks, or longer verbal response times (VRTs). A task such as word recognition in a noisy environment requires an explicit feedback loop and is associated with a greater expenditure of memory resources, thus reducing speech processing speed, since the implicit automatic processing is insufficient [2].

Indirectly, measurement of pupil dilation, reflecting an increase in neuronal activity, can provide information associated with the processing of a degraded acoustic signal. The degradation compels the subject to allocate extra cognitive resources, requiring more cognitive effort [3]. However, in understanding speech, different dissociable processes can be compromised, including verbal working memory and attention-based performance monitoring. The specific resources required will vary depending on the acoustic, linguistic, and cognitive demands of the task, as well as on individual differences in listener skills.

Importantly, to understand the effects of different signal processing technologies available in hearing aids (HAs), and their effects on listening effort, more subtle differences than the usual recognition score are needed. Several authors have shown how amplification can reduce listening effort in adults [4–7]. Hecker et al. [8] emphasize that when various speech communication systems are being evaluated through a conventional intelligibility test, when recognition scores are greater than 90% more sensitive tests are necessary, tests which are able to resolve small differences in how the system is dealing with speech. Pratt [9], who evaluated aircraft communication systems, supports the idea that when the percentage of identification exceeds 90%, the sensitivity of recognition tests can be improved by accessing response times.

During HA fitting, it is usual to test instruments with different technological features or processing strategies so as to address, for example, a speech listening problem in noisy environments or other difficult situation. Although the word recognition score (WRS) is one of the most common ways to obtain information on a person's ability to understand speech, by itself this data is not always sufficient to establish a preference between two devices or strategies, especially if the results are very similar. In such cases, the audiologist will ask the patient to judge listening comfort or how sure they feel about their responses.

Many hearing-health professionals report that during traditional HA evaluations, patients frequently say they have "greater clarity", "more relaxed listening", or "increased listening comfort." The audiologist can often detect a greater or lesser degree of sureness in the answers and, in some cases, perceives a lesser or greater time delay. The idea behind the present work is to measure the response time of the subject in order to quantify the ease of listening during a standard speech recognition test. The proposed explanation for response time differences is based on changes in the cognitive load that different presentation conditions or amplification schemes can cause. Such characteristics can be grouped within the concept of listening effort [10], which is a function of cognitive load. Pichora-Fuller et al. [11] define listening effort as "the deliberate allocation of mental resources to overcome obstacles in the pursuit of objectives when performing a task, with listening effort applied more specifically when the tasks involve listening". Meister et al. [1] suggest that verbal response time (VRT) is a potential marker of cognitive load during conventional speech audiometry. In their work, they compared the VRT obtained in various speech discrimination tasks in noise with the results of a questionnaire on perceived auditory effort, finding that the scales of perceived auditory effort are mainly related to the levels of intelligibility but not so much to the difficulty or ease of listening. Pals et al. [12] experimented with the use of verbal response time measurement to estimate listening effort in single and double-task experiments (visual and auditory) and concluded that a simple auditory task experiment may, as a complement to speech audiometry, perform well as a measure of listening effort.

Such tests might be equally useful when comparing situations with and without auditory equipment, or when varying HA settings, complementing differences in recognition scores. Gatehouse & Gordon [5] proposed the use of response times as a measure of the benefit of amplification. In their work, they pointed out that, in addition to the hearing impairment itself, the perceptual effort that the individual must make to decode a message must also be considered, revealing two types of decoding: bottom-up analysis, which involves reconstruction of the message from the acoustic and phonemic components, and top-down analysis, which can be thought of as the allocation of knowledge about a language's structure to fill in gaps in an incomplete message. When hearing loss or environmental conditions increase recognition difficulty, more top-down processing will be required to maintain performance. These authors claim that it is possible to relate the listening effort to the time elapsed between the message and the response [5]. The present work aims to explore the use of response time during a word recognition test to gather information about listening effort as part of a HA fitting routine.

General methods for obtaining VRT

If one records a word recognition test session, it is possible, using sound editing software, to obtain VRT values by visual inspection of the audio file. Pals et al. [12] analyzed waveforms using an open access sound editor and determined the VRTs from two independent observers to test the reliability of the technique. However, their technique does not work in real-time, requiring post-processing. It also requires the experimenter to visually establish the exact end-point of the presented stimulus and the starting point of the subject's response. Furthermore, if the response signal is recorded along with competing noise, it is virtually impossible to accurately detect, by visual inspection, the moment where the presented speech signal ceases and the response begins. Meister et al. [1] adapted stimulus presentation software (Presentation 16.4) to overcome this problem.

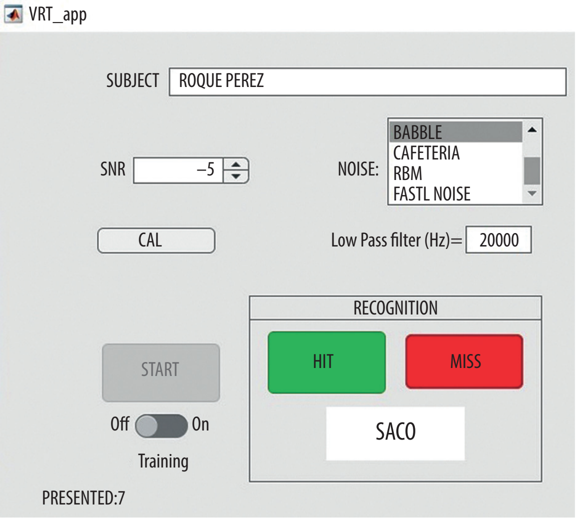
Software for stimulus presentation

For this study, a software program called VRT_app was developed, which runs on a personal computer with an appropriate audio board and the Matlab programming environment. This works as an automatic system that can measure the verbal response times in real time. The test signal can be made to have different levels of audibility, either by applying lowpass filters or adding various types of noise at different SNRs. **Figure 1** shows the graphical user interface of VRT_app which facilitates administration of the test.

Speech material for testing

VRT_app can be used with any type of speech material, such as monosyllabic or disyllabic words, phrases, digits, etc. For the experiments here, 50 two-syllable words were presented, chosen randomly from a set of 700 stored recorded words. The software allows different types of noise to be mixed with the speech signal, with selectable SNR. A variable frequency low-pass filter has also been included, which can be used to restrict speech cues within a certain frequency range, thereby controlling audibility.

In more detail, the software randomly selects words from the 700 words contained in the word lists of Tato et al. [13], Tato & Sarrail [14], and some unpublished lists for children, all recorded and homogenized in terms of presentation level at Mutualidad Argentina de Hipoacúsicos. To generate the stimuli, individual files were created containing each of the words. Each file was "cut" individually, leaving a time of approximately 300 to 400 ms between the beginning of the file and the beginning of the word, and placing the end of the file exactly at the end of the word. This was done by visual and auditory inspection of the selected section, and with great care, since the complete word must be included. When playing the file, the word must sound natural and without any abrupt cut. The program reads each stimulus from file and plays it under the previously specified conditions of noise and filtering. A temporal window opens in which the response given by the subject is recorded. Using this recording, a vector is generated that is processed by the VAD (voice activity detector), which detects the beginning of the subject's response. The time between the beginning of the window and this response (the lapse) is then calculated and stored as a vector (**Figure 2**). Lapse values (in milliseconds) contained in the vector are then sent to an Excel file for analysis and storage.


VRT_app

SUBJECT ROQUE PEREZ

SNR −5

NOISE: BABBLE CAFETERIA RBM FASTL NOISE

Low Pass filter (Hz)= 20000

PRESENTED:7

Figure 1. Graphical user interface of VRT_app for VRT measurements

VAD (voice activity detector)

Several developers of Matlab-compatible software have released add-on modules for voice activity detection (VAD). The objective of a VAD is to detect voice segments and identify the starting and ending points of different words [15]. The system decides the start and end of speech events based on certain criteria. Conventional speech detection techniques are based on two possible approaches, either in the time or frequency domain. Among the first are the cepstral distance method, the zero-crossing rate method, and the temporal energy-based approach. An example of a frequency-based method is the VAD based on spectral energy. Working in the spectral domain allows one to consider the energy distribution across frequencies. An important property of speech is that the human voice preserves spectral components within a certain frequency range, especially at low frequencies. One Matlab-compatible open-access software module for spectral energy VAD, by Chen and colleagues [16], shows good performance even in situations where there is appreciable background noise, and it allows, with small adjustments, to extract the required data.

Comparison between manual and automatic methods

A series of tests were done to compare the results obtained by manual and automatic methods. Four 25-word lists were presented to three different subjects, and recordings of the complete sessions were obtained. A group of three independent trained subjects performed the VRT assessment manually. The difficulties in measuring individual VRTs visually and manually were significant. In real conditions the waveform of speech showed a falling edge, and a word's acoustic characteristics, and the acoustic characteristics of the room where the words were presented, had effects on this falling edge. A room with a longer reverberation time will produce less steep edges. Visual inspection requires a clear definition of how to detect the voice start and end times; here it may help to view the signal in the form of a time series (waveform) or as a spectrogram. The results of

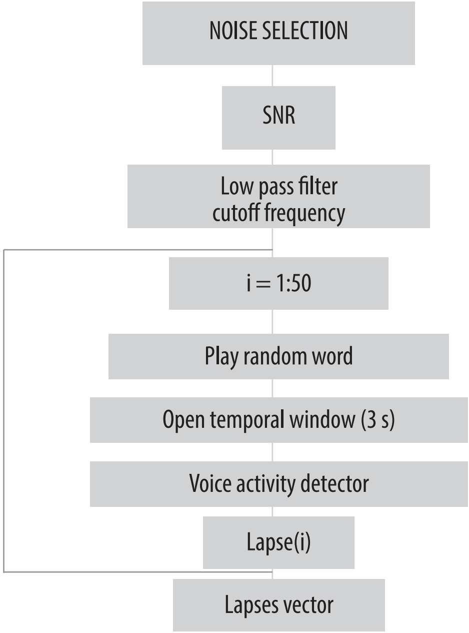


Figure 2. Flowchart of the software program for VRT measurement

measuring an easily assessed signal (a pure 1000 Hz tone with an abrupt falling edge) were compared by both manual and automatic methods, the first with three different observers, obtaining a high degree of correspondence between both measurements ($r(38) = 0.95$, $p < 0.05$). In the case of the automatic method, the system need only detect the beginning of the subject's message. In situations where the test is conducted with background noise, it is difficult, and sometimes impossible, to visually establish the end of the word presented. This makes it preferable to use a type of test where only the start of the response needs to be determined, either manually or automatically. Because it provides immediate results, the automatic method facilitates clinical use, making the test an additional tool in the HA fitting process.

Results of recording and analysis

Once each word is presented, the experimenter receives visual feedback, and the subject repeats the word as recognized. The experimenter registers whether the answer is correct or wrong (or absent) through buttons arranged in the main window. At the end of the presentation of 50 words, an Excel file is created with VRTs and a correct/incorrect registry. Boxplots showing mean and median VRT values, interquartile ranges, and outliers are a good way to quickly access the temporal performance of the subject. As the goal of the measurements is to compare different strategies for HA fittings or cochlear implants, it is convenient to obtain some simple numbers as a basis for comparison. The first and most important parameter is still the word recognition score (WRS). But if this parameter shows small differences between situations, a measure of the central tendency of the VRT (its median) can be added. Data measurements are sometimes skewed or show outliers, so the median is a better choice than the mean. A measure of dispersion is also useful, since early observations indicate that ease of hearing or level of concentration may also affect the dispersion of VRT results. A low dispersion can, in general, be associated with greater ease of listening. Measures of dispersion include the standard deviation (SD) or the interquartile range (IQR). The latter is preferable since it is less affected by outliers and can be easily understood by looking at the height of the box.

To evaluate the system and its possible application, two experiments were carried out to study VRT variations under different listening situations (SNRs and low-pass filtering) in normal hearing people. Also, a third comparison experiment with hearing-impaired subjects using different HA technologies was done to show the potential of these measurements.

Experiment 1: effect of low-pass filtering on VRT in normal hearing subjects

Purpose

This experiment was performed to confirm the effect of reduced audibility on VRTs in subjects with normal hearing. Reduced audibility was achieved by low-pass filtering of the stimulus. Three conditions were tested in random order for each subject: a nonfiltered stimulus and stimuli with low-pass filters of cutoff frequencies 1000 Hz and 1600 Hz.

Method

Eight normal hearing subjects (4 females) with a mean age of 25.6 years (SD = 3.36) participated in the experiment. Before the experiment, normal hearing was verified with pure tone audiometry, and visual inspection of the ear canal was performed. In all the cases, the pure tone air and bone thresholds did not differ by more than 10 dB HL in the frequency range 125 Hz to 8 kHz. For each of two sessions, a complete set of 50 random disyllabic words was presented via a loudspeaker at 1 m (0° azimuth) using the software described earlier and the three filtering conditions. The room where the presentations were made was not acoustically treated, but the noise levels did not exceed those specified by IRAM 4026: 1986 (Argentine standard for audiometric testing) [17].

The speech material was presented at a level judged by each subject as comfortable. A lavalier or handheld microphone was used to capture the subject's responses. Participants were instructed as follows: "You will hear a set of words, some of which will be easier to understand than others. Your task is to repeat them loud and clear as soon as you understand them. You should listen to the entire word before repeating it, to avoid making mistakes." No mention was made to the participants about the speed of the answers. Before the actual experiments, 5 to 10 words were presented to the subjects to familiarize them with the task. The participants completed two test sessions, separated by a period of one week. In each session, the material was presented in 3 conditions: no filter and low-pass filtering at 1600 Hz and 1000 Hz. The order of presentation of the 3 conditions was randomly varied between subjects. Equi-ripple low-pass filters were implemented in Matlab using the FIRPM function included in the Signal Processing Toolbox. The design parameters of the filters were: cutoff frequency = 1000 Hz or 1600 Hz;

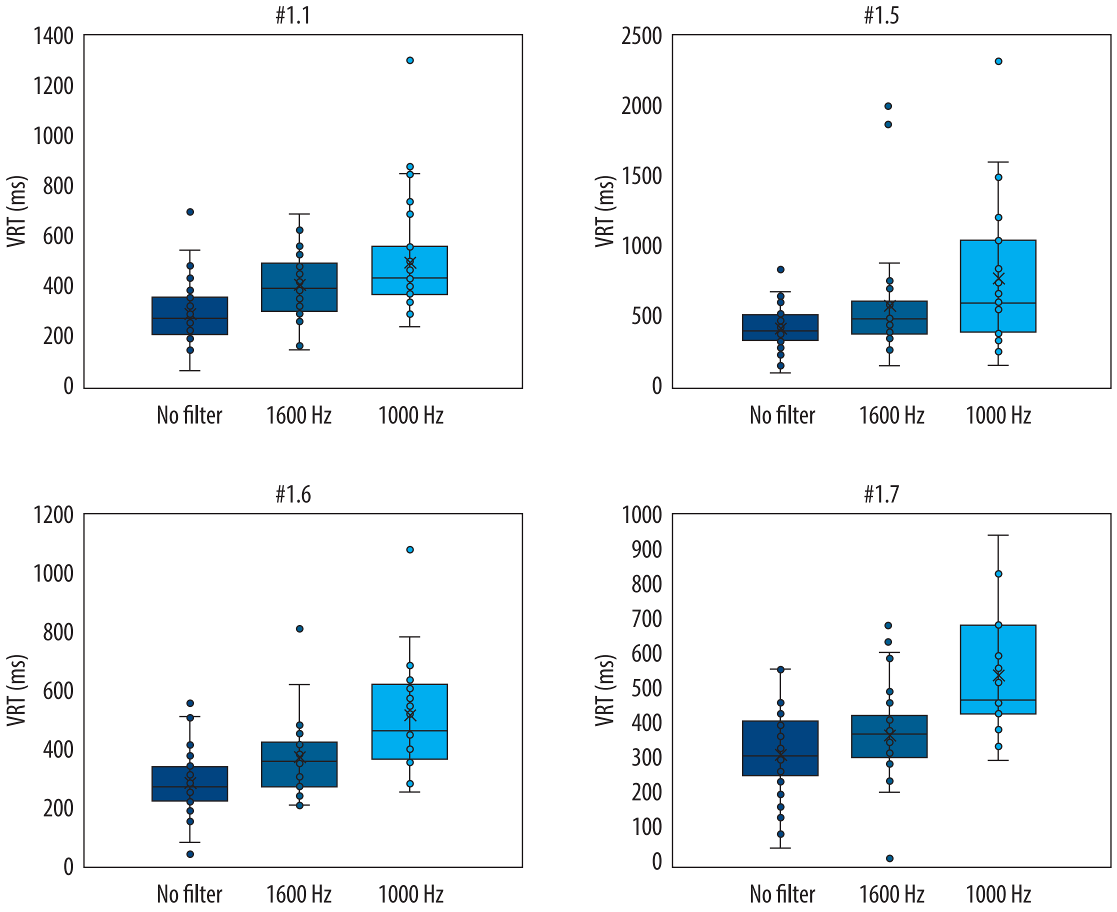


Figure 3. VRTs for correctly identified words under different filter conditions for 4 subjects, from Experiment 1. The boxes define lower and upper quartiles, the horizontal line represents the median, and the cross the mean

stopband frequency = cutoff frequency + 100 Hz; passband ripple = 0.0575; stopband attenuation = 0.0001; density factor = 20.

As with a regular word recognition test, the subject heard the words in noise and repeated what they understood them to be. The experimenter decided if the answer was wrong or right and marked the answer on the main window. For each word the software measured and recorded the elapsed time and the correctness of the repetition.

Results

Individual differences were appreciable, as can be seen in the panels of **Figure 3** for a group of 4 selected subjects. Despite the differences, all showed an increased VRT as the filter became more restrictive, showing that reduced audibility increases response time.

Table 1 shows WRS, median VRT (VRT), and IQR in the filtering conditions for the 8 participants in this experiment. WRSs are in the range 96–100% for the no filter condition, in the range 50–84% for the 1600 Hz filter, and from 20–64% for the 1000 Hz filter.

It is important to note that the VRT values show the response times for words correctly repeated, so reduced audibility has two effects: lower WRS and longer response times for the correctly answered words.

Figure 4 shows the median VRT values for each subject with the filter condition. All the participants showed increased VRT values as the cutoff frequency was lowered. The differences across subjects are shown, with the average for the 8 subjects on the right.

There are large inter-subject differences since each subject uses different criteria which determine response time. Even the same subject showed changes between sessions. As the measure of interest is the change in response time between situations, the within-subject/within-session design and adequate balancing to avoid practice effects could avoid potential problems. The averaged VRTs and WRS for each condition for the 8 participants are shown in **Table 2**. As can be seen, when the cutoff frequency is shifted to lower values, fewer speech cues are available to the listener, and this decreases the WRS and increases the VRT for the repeated words.

Table 1. Recognition rate for words (WRS), median verbal response time (VRT), and interquartile range (IQR) for 8 subjects under three audibility conditions (no filter, and low-pass filters of 1600 and 1000 Hz)

	No filter			1600 Hz			1000 Hz		
	WRS (%)	VRT (ms)	IQR (ms)	WRS (%)	VRT (ms)	IQR (ms)	WRS (%)	VRT (ms)	IQR (ms)
#1.1	100	287.21	144.08	84	403.68	192.11	64	495.80	192.11
#1.2	96	204.73	128.07	72	410.48	128.07	62	480.83	176.10
#1.3	100	550.87	148.08	50	749.89	200.11	42	726.53	216.12
#1.4	96	547.74	108.06	58	676.98	192.11	20	920.54	660.37
#1.5	96	406.72	184.10	52	576.96	232.13	38	767.62	656.37
#1.6	100	284.27	116.07	52	367.00	152.09	42	518.41	256.15
#1.7	96	323.95	156.09	64	372.73	120.07	44	545.86	256.15
#1.8	100	189.19	160.09	72	395.22	124.07	54	477.92	304.17

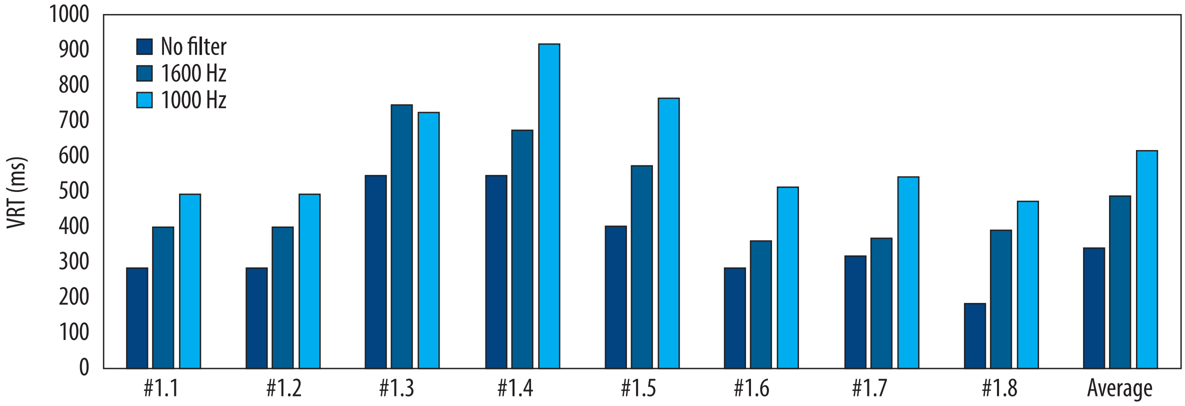


Figure 4. Median VRT for 8 subjects under 3 filter conditions in Experiment 1; average values on the right

Experiment 2: effect of noise on VRT in normal hearing subjects

Purpose

This experiment was performed to confirm the effect of background noise on VRTs in subjects with normal hearing. The participants completed one test session where the speech material was presented in quiet and one with background noise.

Method

A different group of eight normal hearing subjects (4 females) with mean age of 22 years (SD = 3.5 years) participated in the experiment. Normal hearing was verified with pure tone audiometry and visual inspection of the ear canal was performed. Both conditions (quiet and noise) were presented in random order, and the 50 random disyllabic words were taken from a set of 700 words, as explained previously, and competing babble (4-talker babble) was delivered via two loudspeakers at 0° and 180° azimuth. The front loudspeaker delivered the speech signal and the rear loudspeaker delivered the competing noise. The babble involved 3 females and 1 male, as used in previous work [18]. The SNR for this experiment was fixed at 0 dB. Other conditions were the same as in Experiment 1.

Results

The panels in **Figure 5** show the performance of 6 selected subjects who performed the test under both SNR conditions. Differences between the participants are appreciable, although the same pattern of impaired performance is clear in most cases when the situation with competing noise is compared with the quiet situation. Only subject #2.6 showed similar performance under both conditions

Table 3 shows WRS values and VRT (median and interquartile range). These measurements reveal that for some subjects, the WRS difference between quiet and babble could be as small as 2 or 3 words (4 or 6%), but the VRT difference is still appreciable. This fact adds a second parameter to consider when evaluating performance in the recognition task. As can be seen, the changes in VRT between both conditions show a large range of participants. Some of them showed differences of 100% between the two conditions, but subjects #2.6 and #2.8 showed small negative differences, meaning lower median VRT values in the noise condition.

Table 2. Mean values of word recognition score (WRS) and median VRT for 8 subjects under 3 different filter conditions

	No filter	1600 Hz	1000 Hz
WRS (%)	98	66.6	45.5
VRT (ms)	349.34	494.12	616.69

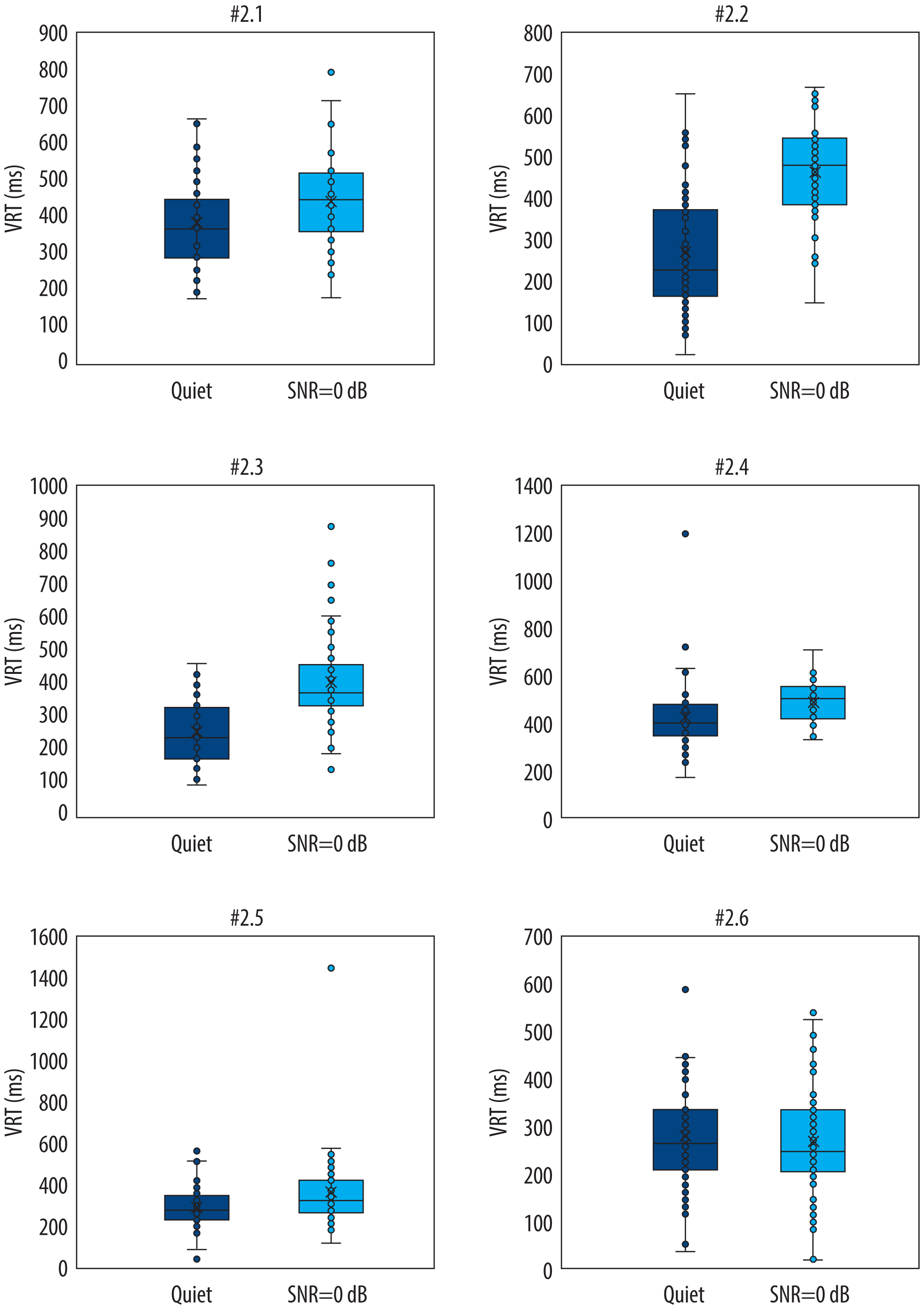


Figure 5. VRT boxplots for 6 subjects in quiet and babble at 0 dB SNR, from Experiment 2

Table 3. Word Recognition Scores (WRS%), median value of VRT, and interquartile range (IQR) for the 8 subjects of Experiment 2 under three different filtering conditions

Subject	Quiet			Babble		
	WRS (%)	VRT (ms)	IQR (ms)	WRS (%)	VRT (ms)	IQR (ms)
#2.1	98	383.26	160.09	88	445.73	160.09
#2.2	92	268.33	212.12	100	465.68	164.09
#2.3	88	260.84	156.09	96	411.16	124.07
#2.4	98	419.53	136.08	92	479.95	136.08
#2.5	100	286.91	116.07	76	363.60	164.09
#2.6	96	277.85	128.07	92	267.65	132.07
#2.7	98	148.35	128.07	92	293.06	116.07
#2.8	98	475.07	96.05	94	465.31	144.08

Experiment 3: within-subject comparison of HA performance in noise based on VRTs and WRSs

The main goal of the present work was to study the possible application of VRT measurement for the evaluation of HA fitting, as suggested by Gatehouse & Gordon [5]. These authors calculated a benefit index based on a comparison of VRT between the aided and unaided conditions, and showed that changes in this index are substantially greater than the conventional index based on simple changes in recognition rate. Also, in cases where the WRS exceeded 85%, or when the difference between two equipment conditions was less than 6%, the VRT values showed significant differences.

Method

Fifteen hearing-impaired subjects (8 females), ages ranging from 32 to 89, median age 61, were tested during HA fitting sessions. The subjects covered a wide range of hearing loss degrees, unaided WRSs, and ages. Since the experimental scheme was a within-subject, within-session comparison of different adjustments of the same HAs, or between different HA models and aided and unaided conditions, the results needed to be analyzed individually. The material and presentation parameters were the same as those used in Experiment 2, with the speech delivered from the loudspeaker and noise from the rear. In most cases, SNR was fixed at 0 dB unless otherwise specified; however, if the noise performance of a patient was very low, the tests were done with a more convenient SNR to achieve better WRS values for comparing VRTs. In all cases, the HAs were calibrated using the manufacturer's rules, with some fine-tuning so as to provide comfort to the patient during a fitting session. In Argentina, where 25-word lists are usually used to verify the performance of a HA, one unrecognized word represents a 4% decrement in the recognition percentage, with 3 words making a 12% difference. Considering the inherent differences in difficulty between lists, and other factors that can change some answers, 12% is a difference that may be within the margin of error and should not be considered as conclusively favoring one situation over another. In marginal cases, examining the temporal dimension can help confirm or rule out these differences.

Isolated words were used as material in these experiments. Gatehouse & Gordon (1990) preferred sentence-based material because the differences in response time seemed greater than when using single words [5]. If the intention of a test is to quantify the degree of top-down processing, sentence-based testing is justified, but in the present work, a word-based approach was used to minimize effects derived from working memory, linguistic context, cultural characteristics, and the like. Our aim was to test access largely to perceptual information. McCreery et al. [19] showed that both working memory and linguistic skills play an important role in speech-in-noise recognition. Subjects with higher vocabulary ability generally had better recognition for sentences in noise, but not for words in noise. Therefore, to standardise the test among different age ranges, working memory skill, and linguistic skill, words were used. Using the minimal linguistic context helps to isolate the results from secondary influences. Because we used a 50-word list in our experiment, each word represented a 2% change in recognition score.

Results

Table 4 shows results for 15 patients that were tested for VRT during a routinary fitting session. The type and configuration of the hearing loss are detailed as well as information regarding the tested conditions. WRS% is the word recognition score, the percentage of correct repeated words. TRVM is the median value of VRT and IQR is the interquartile range (difference between 3rd and 1st quartile), whereas H500 is the percentage of recognized material at 500 ms. Most of the cases show the comparison of two conditions, and a few between three situations. To add the VRT data to the WRS information, the median value is shown in each case as a measure of central tendency. Also, some measure of dispersion is useful as early observations indicate that ease of hearing may also influence the dispersion of VRT results. A lower dispersion could be associated, in general with easy or relaxed listening. Options that can be used to measure dispersion include the standard deviation or the interquartile range. The latter is chosen since it is less affected by outliers. The HAs used are in some cases the own patient HA (usually HA1) and in some other new hearing instruments. They are referenced in **Table 4** by the number of processing channels

Table 4. Data for participants of Experiment 3 comparing the results of WRS and VRT under two or three conditions. Different kinds of hearing loss and equipment are noted (R, right; L, left). VRT is the median value of VRT, IQR is the interquartile range, and H500 is the number of correct words (hits) in 500 ms

Subject	Hearing Loss	Condition	WRS%	VRT (ms)	IQR	H500
#3.1	Mixed severe bilateral	HA1	62	335.68	128.07	0.54
		HA2	72	320.20	156.09	0.66
#3.2	Bilateral conductive moderate	no HA	94	422.52	208.12	0.72
		HA 1 (8 ch)	94	184.82	112.06	0.94
		HA 2 (16 ch)	96	168.29	108.06	0.94
#3.3	Bilateral sensorineural severe	HA1 (8 ch)	88	303.02	124.07	0.76
		HA2 (64 ch)	84	214.69	140.08	0.86
#3.4	R: Sensorineural moderate-severe L: Sensorineural profound	HA1 (64 ch)	68	276.89	248.14	0.64
		HA2 (48 ch)	84	207.75	168.10	0.84
#3.5	Bilateral sensorineural severe	HA1 (8 ch)	44	337.67	308.17	0.38
		HA2 (16 ch)	70	269.35	256.15	0.58
#3.6	Sensorineural bilateral moderate	no HA	52	389.17	252.14	0.44
		HA1 (8 ch)	34	501.01	352.20	0.27
		HA2 (48 ch)	72	233.04	80.05	0.7
#3.7	R: Mild mixed L: Severe mixed	no HA	70	589.61	512.29	0.4
		HA (48 ch)	82	163.32	128.07	0.76
#3.8	R: Sensorineural severe L: Sensorineural profound	HA1 (16 ch)	52	495.07	260.15	0.3
		HA2 (64 ch)	44	331.85	100.06	0.4
#3.9	Sensorineural moderate unilateral	no HA	96	346.77	140.08	0.84
		HA1 (48 ch)	92	317.07	132.07	0.84
#3.10	L: Sensorineural moderate-severe R: profound	no HA	36	860.07	384.22	0.02
		HA1 (16 ch)	46	618.11	320.18	0.16
		HA2 (48 ch)	42	480.29	328.19	0.26
#3.11	L: Mixed moderate R: Mixed moderate-severe	no HA	62	226.73	96.05	0.62
		HA (8 ch)	74	150.77	96.05	0.74
#3.12	R: Mixed mild L: Mixed moderate-severe	HA1 (8 ch)	82	409.05	184.10	0.68
		HA2 (48 ch)	76	334.53	144.08	0.7
#3.13	L: Mixed severe R: profound	only CI	40	521.58	284.16	0.26
		CI+HA (64 ch)	60	376.77	180.10	0.48
#3.14	Mild-moderate bilateral	no HA	56	256.17	136.08	0.52
		HA1	54	158.88	80.05	0.54
		HA2	58	132.51	136.08	0.58
#3.15	Unilateral sensorineural moderate-severe	no HA	88	361.16	361.16	0.82
		HA (48 ch)	96	291.25	291.25	0.94

because this indicator is an easy way to reference technological characteristics, as more processing channels are included on a HA, other characteristics (directional microphones technology, noise suppression, feedback management, compression,) are in general, more advanced.

Figure 6 compares 15 patients under the different tested conditions. Each box shows the range of values between the 1st and 3rd quartile.

Discussion

In this work, we have not investigated relationships between VRTs obtained under different conditions and parameters, such as age, cognitive ability, or self-rating of hearing difficulty. Instead, our experiments were designed to compare combinations of listening difficulty and hearing equipment. To replace the subjective judgment of the audiologist when modifying an amplification setting, we sought to obtain an objective measure of the degree of confidence, one which involves the response time, VRT.

In the experiments with normal hearing people, VRT was found to be sensitive to reductions in audibility induced by low-pass filtering and adding noise. The word lists used to construct the material were based on various author lists currently in use in Argentina. A study of how our particular results vary with word difficulty has not yet been done, but it may be incorporated in future research. The main idea of this work was to demonstrate the basic technique and test that it works. Of course, refining the quality and homogeneity of the test method will make it more reliable.

Although the magnitude or rate of the variations obtained by filtering the signal or adding background noise have not been studied, it is clear that shorter VRTs are obtained when the conditions are less difficult. In the experiments with hearing-impaired patients, it was possible to compare the use of different technologies, especially noise reduction algorithms. With mild or moderate hearing loss and high WRS levels, measuring VRT allowed us to gauge the practical advantage or convenience of different HA fitting strategies. In cases where speech recognition was good, differences in VRT reflect higher or lower listening effort. As Gatehouse & Gordon pointed out [5], an index of amplification benefit based on VRT can offer advantages over the traditional WRS measure in cases where there is a ceiling effect or the difference between test conditions is within the test–retest reliability range.

Looking closer at **Figure 6** and **Table 4**, some cases can be discussed in more depth. Comparing boxplots for subject #3.2 (bilateral conductive moderate hearing loss), it is easy to see the advantage the HAs confers compared to the non-aided condition. Although the recognition score is high and the differences between the three situations are small, the benefit in terms of VRT is clear. Other methods for deciding the suitability of equipment, based solely on an improvement in recognition scores, would not have yielded a convincing result (given the high scores and small differences between the situations). A difference in recognition score of just 2% between the cases (94 and 96%, just one word) is not reliable enough to decide the question, but with VRT it is clear that the two HAs provide reductions of the order of 50% (from a response time of 368 ms to 176 or 160 ms).

Patient #3.3 showed similar recognition scores with both HAs (a difference of just 4%). However, the use of HA2, a 16-channel device with noise management and improved directional capabilities, brings about a reduction of VRT close to 40 ms compared with the previous 8-channel HA1. Although the difference in the median VRT between both conditions is not large, the boxplot in **Figure 6** shows that 50% of the VRT values are below those corresponding to the old HA, revealing the higher performance advantage of HA2 over HA1.

In cases where the recognition score shows greater improvements, as with patient #3.7, the VRT improvements are also greater. In this case, the H500 index showed that, with no HA, the subject was able to correctly repeat only 40% of the material within 500 ms, compared with 76% in the aided condition. Case #3.10 also showed a big decrease in VRT as the technology employed was improved, even though recognition scores remained about the same. With the best technology (48 channels and better noise management), the response time falls from more than 700 ms to almost 300 ms, even though the change in recognition score was only 6% (3 words) compared with the unaided condition. This finding corresponds with the patient's subjective perception of more relaxed and clearer listening (despite the recognition scores not reflecting much improvement).

Case #3.13 is a CI user who tried bimodal equipment. Here the comparison is between the CI-only and the CI+HA condition. Both figures, WRS and VRT, improved in the bimodal condition. In this case, the increase in the recognition score alone justifies the adoption of bimodal equipment, although the significant decrease in the VRT value confirms the convenience of adding a HA. Comparing the boxplots, the global improvement in response times is clear (which can be seen both in the height of the box, with less dispersion, and in comparing the extreme values).

In some cases, like #3.1, there is no appreciable difference between the two situations in terms of VRT, although there is a difference of 10% (5 words) in the recognition scores. Here, it seems the VRT was not a parameter that helped decide whether the old or new hearing instrument was better. By way of contrast, in cases #3.3, #3.8, and #3.10, the VRT measurements could help decide which was the better HA configuration. Case # 3.3, a bilateral severe hearing loss, had slightly lower recognition for the 64-channel HA than the 8-channel instrument (84% vs. 88%) but faster response times (224 ms vs. 264 ms). In certain cases, the patient's judgment about relaxed listening is enough, but there are situations where the patient's preference is not so clear and the VRT measurement can be useful. Similar results are obtained in the case labeled #3.8, with better response time but a lower recognition rate. This might be solved with some fine-tuning of the 64-channel instrument so as to improve recognition scores. Case #3.10 showed better recognition rates in the unaided condition, but both the patient's judgment and the VRT showed an improvement with the 48-channel HA. This was a bilateral hearing loss, moderate-severe in one ear and profound in the other. The results for the aided condition are not good with any of the HAs, but the second did show an improvement in VRT.

Almost all the cases in Experiment 3 were presented at 0 dB SNR. The exception was #3.13, with an SNR of + 5 dB. To obtain better WRS measures, it is possible to increase the SNR, and in some cases of very poor patient performance in noise, it is possible to make a VRT comparison without any noise at all, with VRTs now showing differences between two adjustments or fittings. It is suggested

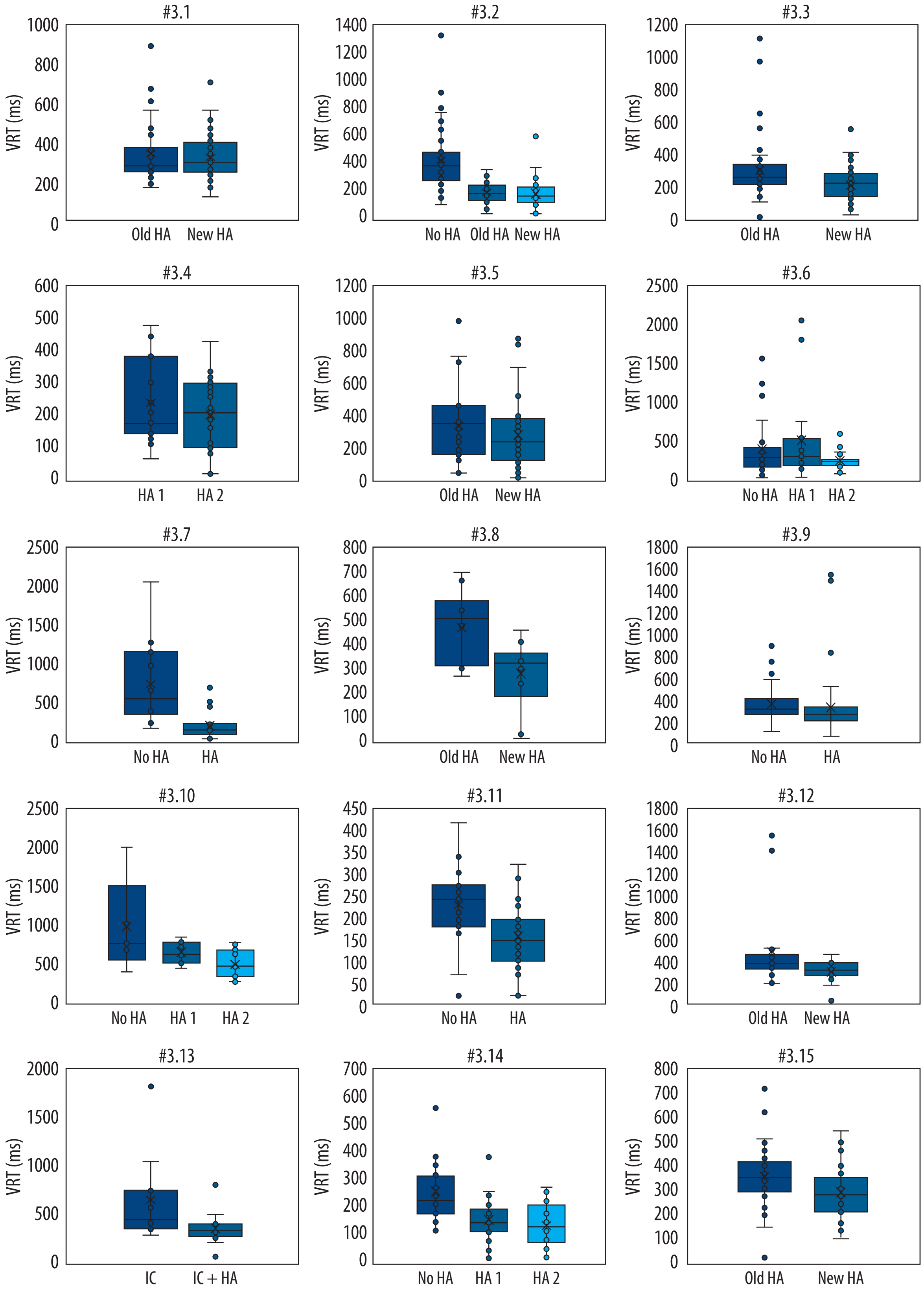


Figure 6. VRT boxplots for 15 hearing-impaired subjects in Experiment 3. The boxes for each patient show VRT values for the conditions described

that just comparing median VRT values or examining box plots can be a good way of gauging ease of listening.

In recent work, Carolan et al. [20] pointed out that behavioral measures, such as VRTs, may indicate when an individual's resource capacity is under strain or exceeded, allowing inferences to be made about listening effort. Increased effort may manifest as slower response times due to longer processing time, but they flagged that this needed to be considered carefully because the relationship between listening effort and response time depends on the experimental context.

Before applying VRT measurements to older adults as an indicator of cognitive load, it should first be recognised that this group presents worse results than younger adults in recognition tasks [21]. Among the reasons for their difficulties in adverse environments is the impairment in peripheral hearing that typically accompanies age, but also accompanying deficits in sound segregation, less ability to distinguish voices, and cognitive problems in general. Response times can be aggravated by these age-related problems, since the recruitment of greater cognitive energy to compensate for deficits in hearing is perceived by the subject as greater listening effort, and this can be reflected in longer response times. The use of response times may well be complementary to the patient's own evaluation of perceived effort.

As pointed out by Helfer et al. [22], as subjects get older they are more susceptible to difficulties in noisy environments, especially when the competitive noise is speech. The use of questionnaires such as the SSQ can provide tools to measure the self-perceived level of listening difficulty, and our proposal to measure response times is part of a search for measurable parameters that help assess hearing difficulty. VRT is not intended to replace procedures but to complement them. The addition of a VRT measurement does not require a separate test to be done.

Regarding the risks of using this type of indicator of cognitive load on older adults, they are no different from those of any conventional speech recognition test used for HA fitting. However, it must be remembered that response times inherently vary due to many subjective factors: comfort of the subject at the time of the test, their emotional situation, cognitive awareness, working memory, peripheral hearing, and so on, all of which contribute to the large differences found between subjects. The statistical nature of measuring a median VRT involves multiple factors, both within-subject and within-session, and will inevitably lead to a spread in VRTs obtained in the same subject. Nevertheless, the use of a software program that automatically measures the response times and shows the box plots on screen could be a valuable tool to confirm or reconsider the results of recognition scores.

Conclusions

This work has shown how the use of verbal response times in word recognition tasks introduces a second valuable parameter that can provide more detailed information about listening effort than a simple subjective response from the patient. VRT is objective, and quantifies aspects of performance not found in the recognition score. Although today's HA fitting techniques rely increasingly on the use of prescriptive formulas and verification with real ear measurements, eventually it ends up as a subjective test of HA performance. In many cases the reason is economic, since few audiologists have access to sophisticated equipment, but also because the benefit itself is subjective, since ultimately the subject themselves has to decide whether the device is helpful or not.

A test of speech recognition in noise confronts the subject with a complex situation that is often difficult to decide on the spot. The benefits of a noise reduction algorithm or other improvements are not always easily verifiable in the office during one fitting session. The concepts of listening ease or listening effort have been gaining increasing importance, and it is reasonable to think that by improving these technological aspects, the patient will benefit in terms of higher quality hearing. The first step is to achieve an adequate word WRS, following which the VRT can be used as a tool that allows finer comparisons between settings to be made. In cases where the need for a HA is moot (unilateral or slight losses), a comparison of VRTs between aided and unaided conditions can help decide the issue. A quick visual inspection of box plots can help appreciate how global response times react to a change in fitting, and this could be useful in the clinic.

When comparing fitting strategies, response time adds a new temporal dimension to the evaluation. Including an assessment of VRT in a speech recognition test has the advantage of providing a more objective marker of the ease of listening without adding time to the test itself. Implementation of a system that records and processes response times, at the same time as the WRS is being assessed, can provide valuable information for fine-tuning the best HA setting.

Acknowledgments

The authors thank Carla Minniti, Aud. Lic, for her supporting work.

References

1. Meister H, Rählmann S, Lemke U, Besser J. Verbal response times as a potential indicator of cognitive load during conventional speech audiometry with matrix sentences. Trends Hear, 2018; 22: 2331216518793255.
2. Rönnberg J, Lunner T, Zekveld A, Sörqvist P, Danielsson H, Lyxell B, et al. The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. Front Syst Neurosci, 2013; 7: 31.
3. Brennan MA, Lewis D, McCreery R, Kopun J, Alexander JM. Listening effort and speech recognition with frequency compression amplification for children and adults with hearing loss. J Am Acad Audiol, 2017; 28(9): 823–37.
4. Downs, DW, Crum MA. Processing demands during auditory learning under degraded listening conditions. J Speech Hear Res, 1978; 21(4): 702–14.

5. Gatehouse S, Gordon J. Response times to speech stimuli as measures of benefit from amplification. Br J Audiol, 1990; 24(1): 63–8.
6. Humes LE. Dimensions of hearing aid outcome. J Am Acad Audiol, 1999; 10(1): 26–39.
7. Picou EM, Ricketts TA. The effect of changing the secondary task in dual-task paradigms for measuring listening effort. Ear Hear, 2014; 35(6): 611–22.
8. Hecker MH, Stevens KN, Williams CE. Measurements of reaction time in intelligibility tests. J Acoust Soc Am, 1966; 39(6): 1188–9.
9. Pratt RL. On the use of reaction time as a measure of intelligibility. Br J Audiol, 1981; 15(4): 253–5.
10. Peelle JE. Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. Ear Hear, 2018; 39(2): 204–14.
11. Pichora-Fuller MK, Kramer SE, Eckert MA, Edwards B, Hornsby BWY, HUmes LE, et al. Hearing impairment and cognitive energy: the Framework for Understanding Effortful Listening (FUEL). Ear Hear, 2016; 37: 5S–27S.
12. Pals C, Sarampalis A, van Rijn H, Başkent D. Validation of a simple response-time measure of listening effort. J Acoust Soc Am, 2015; 138(3): EL187–92.
13. Tato JM, Sarrail EC. Determinacin de las características del españolen Buenos Aires (Determination of the characteristics of Spanish in Buenos Aires). Ann Otorrinolaringol Ibero Am, 1974; 1(5): 15–29.
14. Tato JM, Lorente Sanjurjo F, Bello J, Tato JM. Características acústicas de nuestro idioma (Acoustic characteristics of our language). Revista de la Federación Argentina de Sociedades de Otorrinolaringología, 2004; 1948–67.
15. Park JS, Yoon JS, Seo YH, Jang GJ. Spectral energy-based voice activity detection for real-time voice interface. J Theor Appl Inf Tech, 2017; 95(17): 4304–12.
16. Chen YW. Voice activity detection by spectral energy. https://github.com/JarvusChen/MATLAB-Voice-Activity-Detection-by-Spectral-Energy [Available 2021.07.25].
17. IRAM 4026: 1986. Cabinas audiométricas (Audiometric booths). Instituto Argentino de Normalización y Certificación. https://catalogo.iram.org.ar/#/normas/detalles/4626
18. Cristiani HE, Serra V, Guinguis M. Development of a quick speech-in-noise test in “Rioplatense” Spanish, based on Quick – SIN®. J Phonet Audiol, 2020; 6(1): 145.
19. McCreery RW, Spratford M, Kirby B, Brennan M. Individual differences in language and working memory affect children’s speech recognition in noise. Int J Audiol, 2017, 56(5): 306–15.
20. Carolan PJ, Heinrich A, Munro KJ, Millman RE. Quantifying the effects of motivation on listening effort: a systematic review and meta-analysis. OSF Preprints, 2021.
21. Helfer KS, Freyman RL. Stimulus and listener factors affecting age-related changes in competing speech perception. J Acoust Soc Am, 2014; 136(2): 748–59.
22. Helfer KS, Merchant GR, Wasiuk PA. Age-related changes in objective and subjective speech perception in complex listening environments. J Speech Lang Hear Res, 2017; 60: 3009–18.