G Model
ANORL-507; No. of Pages 6
ARTICLE IN PRESS

European Annals of Otorhinolaryngology, Head and Neck diseases xxx (2016) xxx–xxx

Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com/en

Original article

Validation of the French-language version of the OTOSPEECH automated scoring software package for speech audiometry

F. Venail[a,b,*], E. Legris[b], B. Vaerenberg[c], J.-L. Puel[b], P.J. Govaerts[c], J.C. Ceccato[c]

[a] *Service ORL, université Montpellier, CHU Gui-de-Chauliac, 80, avenue Augustin-Fliche, 34295 Montpellier cedex, France*
[b] *Plateforme d'audiologie I-PAudioM, institut des neurosciences de Montpellier, Inserm U1051, 34091 Montpellier cedex, France*
[c] *The Eargroup, Herentalsebaan 75, 2100 Antwerp-Deurne, Belgium*

ARTICLE INFO

Keywords:
Speech audiometry
Automated scoring
Validation study

ABSTRACT

Objectives: To validate a novel speech audiometry method using customized self-voice recorded word lists with automated scoring.
Patients and methods: The self-voice effect was investigated by comparing results with prerecorded or self-recorded CVC (consonant-vowel-consonant) word lists. Then customized lists of 3-phoneme words were drawn up using the OTOSPEECH software package, and their scores were compared to those for reference lists. Finally, the customized list scores were compared on automated (Dynamic Time Warping [DTW]) versus manual scoring.
Results: Self-voice did not change scores for perception of CVC words at 10, 20 and 30 dB (ANOVA > 0.05). Scores obtained with pre-recorded and self-recorded lists correlated ($n = 10$, $R^2 = 0.76$, $P < 0.01$). Customized list scores correlated strongly with the reference cochlear lists of Lafon in normal-hearing ($n = 77$, $R^2 = 0.83$, $P < 0.001$) and hearing-impaired populations ($n = 13$, $R^2 = 0.89$, $P < 0.001$). Results on the automated and manual scoring methods correlated in both populations ($n = 77$, $R^2 = 0.71$, $P < 0.01$; and $n = 13$, $R^2 = 0.76$, $P < 0.01$, respectively), with DTW scores ranging from 24.17 to 53.24.
Conclusions: Automated scoring of customized self-voice recorded lists for speech audiometry displayed results similar to conventional audiometric techniques.

1. Introduction

Speech audiometry is an everyday practice in audiology, not only to diagnose hearing loss but also to assess auditory rehabilitation (conventional or implanted hearing aid, cochlear implant or speech therapy).

Implementation involves the subject hearing then repeating phonemes presented at varying intensity. The examiner judges whether the phonemes have been repeated correctly, and scores the subject's performance (usually as a percentage).

Speech audiometry performance depends not only on auditory acuity and presentation intensity but also on the type of vocal material presented [1]. The stronger the semantic content of the material (e.g., real words rather than nonsense words or logatoms), the more understanding calls upon the subject's lexical field [2].

Thus speech audiometry assessment runs up against insurmountable obstacles if subjects are not being tested in their native language, if their lexical field is limited or if a strong regional accent makes it difficult for the examiner to understand the repeated phonemes [3]. Assessment may further be biased in the context of rehabilitation programs, where iterative presentation of the same word-lists induces learning bias over successive sessions, artifactually boosting the subject's scores.

The present study assessed the OTOSPEECH software package (Eargroup, Antwerp, Belgium), which uses customized word-lists drawn from the individual subject's own lexical field, with automated scoring. Briefly, word-lists are taken from the subject's everyday lexical field according to the occurrence of phonemes in his or her native language (including dialects) [4], then recorded with the subject's own voice, thereby allowing for individual acoustic and articulatory parameters and accent. They are then presented at varying intensity, just as in conventional speech audiometry. The words as repeated by the subject are then compared against the recorded words, using an algorithm to determine automatically the number of correctly repeated phonemes.

This semi-automatic speech audiometry procedure was assessed by first comparing subjects' performance when repeating logatoms recorded in their own voice or by another speaker.

* Corresponding author. Tel.: +33 4 67 33 68 00; fax: +33 4 67 33 67 28.
E-mail address: frederic.venail@inserm.fr (F. Venail).

http://dx.doi.org/10.1016/j.anorl.2016.01.001

Please cite this article in press as: Venail F, et al. Validation of the French-language version of the OTOSPEECH automated scoring software package for speech audiometry. European Annals of Otorhinolaryngology, Head and Neck diseases (2016), http://dx.doi.org/10.1016/j.anorl.2016.01.001

Performance was next compared between customized word-lists and calibrated lists, both self-recorded. Finally, automatic scoring was compared to against the examiner's manual scoring.

2. Materials and methods

2.1. Subjects

The study population comprised native French-speaking adults able to read easily from a computer screen placed at a comfortable distance. Educational levels were at least middle-school, and speech intelligibility (Speech Intelligibility Rating [SIR]) was excellent (SIR 1).

Pure-tone audiometric hearing thresholds were checked after ruling out external or middle ear pathology on otoscopic examination. Subjects with conduction or perceptual hearing loss were excluded. All measurements concerned acoustic stimuli in the right ear.

Two subgroups were studied: normal-hearing and hearing-impaired according to ISO 7029.

The study was considered as comprising routine care, and did not require institutional review board approval.

2.2. Material

The study was conducted at the I-PAudioM audiology platform (INSERM U1051), Montpellier (France). Testing was performed in soundproof booths, using a calibrated Affinity system (Interacoustics, Denmark). Audiometric thresholds were studied for each ear separately, using TDH-39 headphones, and conduction hearing loss was ruled out by bone conduction measurements using a B-71 vibrator.

The computer linked to the Affinity system was connected up to a USB sound card (Aureon 7.1) and microphone to record the subject's voice. Visualization equipment comprised two screens: one in the booth, visible to the subject, and one outside, visible to the examiner.

The OTOSPEECH software is part of the A§E suite (Otoconsult, Antwerp, Belgium) and was installed on the computer connected up to the measurement system; the output intensity of the word-lists emitted by the program was calibrated using a sonometer.

2.3. Voice recording and results analysis on OTOSPEECH software

The OTOSPEECH software enabled use of prerecorded word-lists pronounced by a speaker or self-recorded lists read by the subject from a text. Customized lists of 3-phoneme CVC (consonant-vowel-consonant) words were taken by the software from a text of more than 300 words chosen by the subject in a field familiar to him or her. The words were then recorded by the subject, after the software had checked the audio quality of the recording set-up. Intensity of each word in the lists was calibrated automatically by the software. The words were then saved for future use. In this way, subjects created their own speech audiometry word-list, consisting of familiar vocabulary and recorded by their own voice.

In the test phase, OTOSPEECH compared each word as repeated by the subject to the source-word recorded by the subject, on acoustic comparison using a dynamic time warping (DTW) algorithm [5]. Unlike other speech-recognition methods, DTW requires no phoneme data-base for comparison. The algorithm extracts indices (Mel-Frequency Cepstral Coefficients [MFCC] [6,7]) and, after normalization, compares the Euclidean distance between them, in the form of vectors calculated for the target signal (recording) and test signal (repetition): the shorter the distance, the closer the target and test signals: i.e., the more faithful the repetition. This type of analysis limits intra-subject variability and requires no learning phase. Implementation of DTW distance was validated in Dutch in a previous study [8]; the DTW score, after polynomial transformation to express it as a value between 0 and 100%, showed the same psychometric properties as a speech audiometry curve using the Bruges CVC logatom test (see [8] for details).

In the light of these results, we did not perform the transformation but, in subsequent analyses, compared DTW scores and correct repetition scores after manual validation.

In parallel, the self-voice effect was studied using CVC logatoms in a subgroup of normal-hearing subjects. Word-lists were recorded using Adobe Audition CS5 software (Adobe Corp., San Jose, CA, USA) and intensity was normalized using Matlab software (Mathworks, Natick, MA, USA).

2.4. Assessment of the self-voice effect in speech audiometry

The impact of using the subject's own voice in speech audiometry was assessed in 2 sessions, in part of the normal-hearing population:

- in the first session, 4 prerecorded (neutral male voice) lists of 17 CVC Dodelé logatoms each [9] were delivered at 5, 10, 20 and 30 dB HL and correct repetition was scored manually;
- at end of session, the 4 lists were recorded by the subject and saved for session 2;
- in the second session, 2 weeks after self-voice recording, the lists were delivered at the same intensities as in session 1 (5, 10, 20 and 30 dB HL) and scored manually.

2.5. OTOSPEECH validation

The OTOSPEECH software was assessed in 2 sessions.

In the first session, 2 Lafon cochlear lists of 17 3-phoneme words each [10] were recorded by the subject and saved to the software. Also, 2 customized lists of 24 3-phoneme words each were put together from a text familiar to the subject.

The second session was held at least 2 weeks later, to limit memory bias. Speech audiometry was performed using the 2 Lafon lists and 2 customized lists, all self-recorded. Lists were presented at 2 intensity levels: one judged week and the other comfortable by the subject; levels thus varied between subjects. Instead of reporting the speech perception scores for fixed intensities, we rather reported the results for two subjective intensities (low and high), chosen by the test subject according to his or her hearing threshold, to be able to show the variations in scores between subjects. The number of correctly repeated phonemes was scored manually for both types of list, and DTW scores were calculated automatically for the customized lists only.

2.6. Statistical analysis

For logatoms, raw values were compared on 2-factor ANOVA ($P < 0.05$ significance threshold) to discern effects of intensity (5, 10, 20 and 30 dB HL) and list type (pre- versus self-recorded).

Correlations between manually and automatically-scored Lafon and customized lists (percentage correctly repeated phonemes) were assessed on Pearson correlation test ($P < 0.05$ significance threshold).

Finally, the correlation between manual scoring (percentage correctly repeated phonemes) and automated scoring of customized lists (DTW score) was assessed on Pearson correlation test ($P < 0.05$ significance threshold).

Please cite this article in press as: Venail F, et al. Validation of the French-language version of the OTOSPEECH automated scoring software package for speech audiometry. European Annals of Otorhinolaryngology, Head and Neck diseases (2016), http://dx.doi.org/10.1016/j.anorl.2016.01.001

Table 1
Pure-tone audiometry thresholds of normal-hearing and hearing-impaired subjects.

	Frequency (Hz)					
	250	500	1000	2000	4000	8000
dB HL threshold ± SD						
Normal-hearing (n = 77)	6.2 ± 5.8	4.3 ± 4.3	4.5 ± 4.8	0 ± 5.2	2.6 ± 2.4	7.5 ± 6.9
Hearing-impaired(n = 13)	16.8 ± 6.7	14.1 ± 5.5	17.3 ± 5.8	23.6 ± 7.0	43.6 ± 7.6	49.5 ± 9.9

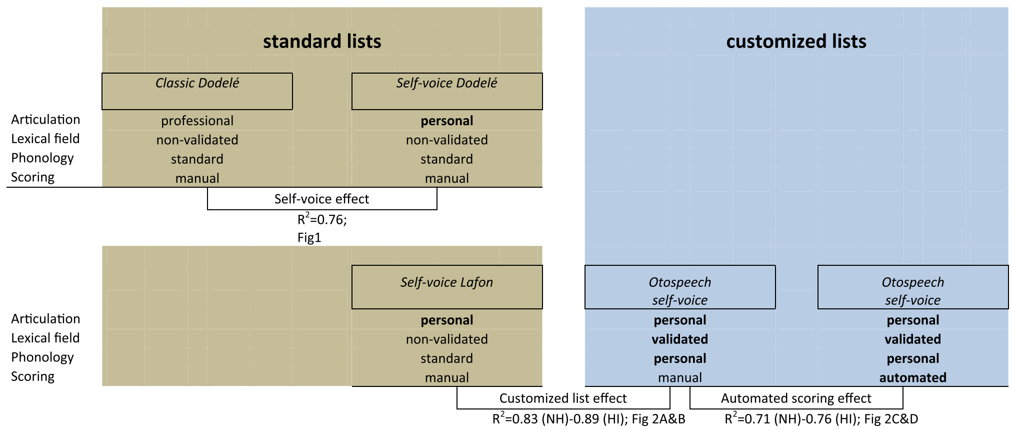

Fig. 1. Overview of procedures, and correlations between study tests. Comparison of standard and customized speech audiology lists, with R^2 correlation coefficients, in the normal-hearing (NH) and hearing-impaired (HI) populations. For each type of test, parameters comprised: recording modality (self-voice or other-voice), whether the words were known to come under the subject's lexical field (validated or non-validated), whether phonologic parameters such as rhythm, intonation or accent were standardized or customized (i.e., self-voice or other-voice), and whether scoring was manual or automated.

3. Results

3.1. Study population

The study population comprised 53 women and 37 men, with a mean age of 31 years (range, 19–67 years). Seventy-seven were normal-hearing (ISO 7029 standard) and 13 hearing-impaired (Table 1).

All were tested to assess the impact of using a customized versus standardized word-list and the correlation between manual and automated scoring of speech audiometry.

Fig. 1 summarizes the study design and correlation coefficients (R^2) between the various test conditions.

3.2. Self-voice effect in speech audiometry

The number of correct phoneme repetitions on the Dodelé logatom test at different intensities was compared between pre- and self-recorded lists in 24 of the 77 normal-hearing subjects.

There was no difference in performance, whatever the intensity, between the self-recorded and the standard list (n = 24, ANOVA, $P > 0.05$) (Fig. 2A). Fig. 2B shows a very good correlation between scores with the two types of recording ($R^2 = 0.76$, $P < 0.01$), although with a slight non-significant advantage for self-recorded logatoms (intercept of 10.2%). However, wider scatter at low intensities reflects greater variation in results between the two test conditions when scores were low.

It can consequently be concluded that using the subject's own voice did not alter logatom comprehension performance in the normal-hearing population.

3.3. Speech audiometry performance using standardized versus customized lists

In the normal-hearing population (Fig. 3A), there was a highly significant correlation ($R^2 = 0.83$, $P < 0.001$) between correct phonemes repetition scores on the Lafon and customized lists; difference in scores was minimal (−2.73%).

In hearing-impaired subjects (Fig. 3B), there was also an excellent correlation between the two types of list ($R^2 = 0.89$, $P < 0.001$), with apparently a slight advantage for the standardized lists (intercept of 16.3%). However, as low scores were less frequent in this population, calculation of the intercept may be inaccurate.

3.4. Comparison between manual and automated scoring in speech audiometry

Manual scoring of correct phoneme repetition on customized lists was compared with automated DTW scoring in the normal-hearing and hearing-impaired groups.

In the normal-hearing population (Fig. 3C), DTW scores ranged between 24.17 (good concordance between recorded and repeated word) and 53.24 (poor concordance). There was a very clear negative correlation between manual and automated scoring of phonemes ($R^2 = 0.71$, $P < 0.01$).

Results in the hearing-impaired population were similar (Fig. 3D), with DTW scores ranging between 27.49 and 50.77. Here again a significant negative linear correlation was found ($R^2 = 0.76$, $P < 0.01$).

4. Discussion

A software package enabling the creation of customized word-lists with automated scoring is of especial interest in the follow-up of hearing loss. There is no limit to the number of lists that can be created, thus getting round the problem of learning bias, and the lists can be administered as self-assessment, being recorded by the subject and scored by the software. This approach may be particularly suited to iterative assessment of hearing-aid or cochlear implant bearers.

The present study assessed the impact of both self-voice recording and customized lists in speech audiometry.

Please cite this article in press as: Venail F, et al. Validation of the French-language version of the OTOSPEECH automated scoring software package for speech audiometry. European Annals of Otorhinolaryngology, Head and Neck diseases (2016), http://dx.doi.org/10.1016/j.anorl.2016.01.001

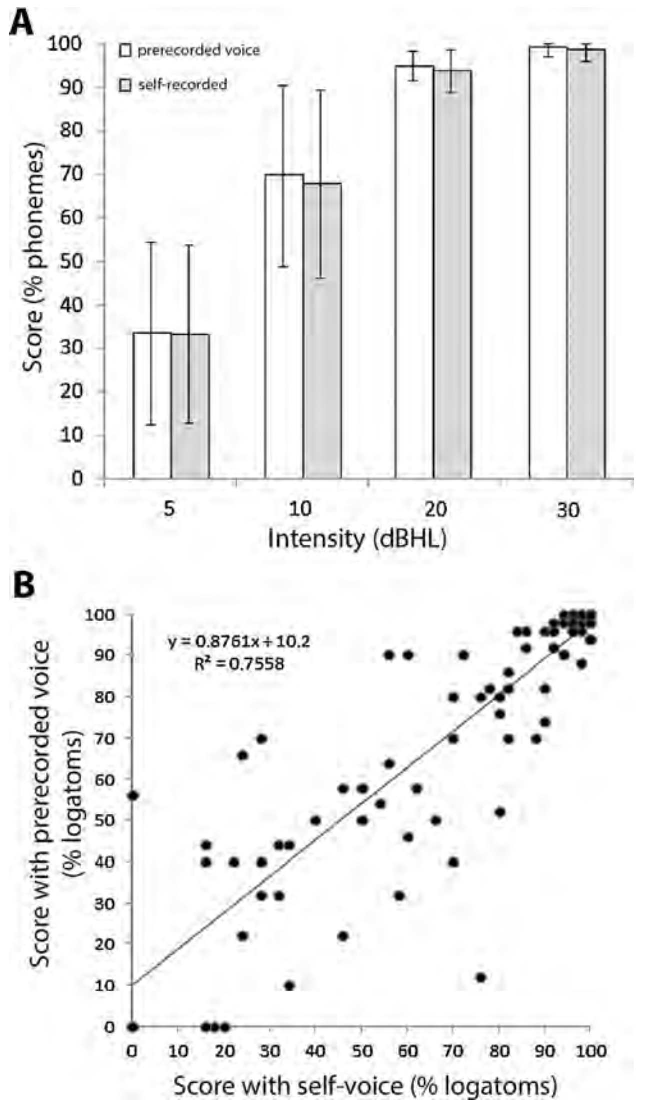

Fig. 2. Comparison of scores with prerecorded and self-recorded logatom lists. A. Percentage correctly repeated phonemes at 10, 20 and 30 dB HL with standard deviation, for prerecorded (white) and self-recorded (gray) lists. There were no significant differences, whatever the intensity. B. There was a significant correlation between scores with prerecorded and self-recorded logatoms. Scatter was greater for low scores, corresponding to low intensity presentation.

Self-voice recording of speech audiometry word-lists gets around the problem of differences in gender, age or regional accent between subject and speaker. In normal-hearing subjects, most studies agree that intelligibility is independent of the speaker's gender [11,12], in both silence and noise [13], so long as the fundamental frequency (F0) of the voice is well perceived [14]. In hearing-impaired subjects or vocoder simulation of hearing loss [15], female voices seem to be less well understood than male voices; but this effect concerns fricative consonants [16] and not vowels, so that CVC logatom lists are more strongly affected than meaningful words, for which the subject can draw on his or her lexical resources to choose words resembling what is heard and thus at least partially compensating for deficient intelligibility [2].

Age differences between speaker and test subject may also induce variation. Zäske et al. [17] reported longer adaptation time in young subjects hearing a voice that had been "aged" by morphing, compared to a coeval voice.

Likewise, differences in regional accent may cause problems: exposure to a new accent has been shown to require a time of adaptation for phonological processing of speech [18], and may rely on different central auditory information processing channels [19].

In the present study, independently of age and for varying degrees of hearing loss, self-recorded logatom lists gave results equivalent to those with prerecorded lists. However, results were more variable when scores were low: i.e., at low presentation intensity. This is because lists are often articulated by a voice professional, speaking especially slowly, so that the conditions are remote from those of real life. Such customized lists could therefore not be brought closer to real life conditions without impairing the quality of the speech audiometry results.

If the speaker's gender, age and accent have an impact, however limited, on speech intelligibility, the subject's recognition of his or her own voice may also influence results. The ability to recognize one's own voice emerges at the age of 4 years [20], even in case of dysphasia. This neurophysiological process may be related to the development of short-term working memory, with a different underlying mechanism from that of individual voice recognition in multi-speaker conversation [21]. However, does this process provide an advantage for speech intelligibility? Using one's own voice may enhance understanding by increasing congruence between visual and auditory information [22]; but the test conditions in speech audiometry involve only auditory input, and hearing one's own voice would provide an advantage only in the particular situation where speech perception is possible only as of the 3rd vocal formant [23], in the unusual case of low-frequency hearing loss with threshold degradation below 2000–2500 Hz.

To sum up, in line with the present results, self-voice recording of speech audiometry lists seems neither to improve nor impair intelligibility, and can thus reasonably be used for recording customized lists.

If self-recording and customized lists can be used in speech audiometry without affecting results, it seems reasonable also to use automated scoring, based on correspondence between sounds heard and repeated with the same voice.

The dynamic time warping (DTW) algorithm is based on analysis of mel-frequency cepstral coefficient (MFCC) vectors measured on phonemes that can be normalized to eliminate the impact of inter-subject variation on speech recognition [6], thus getting round issues of gender or accent. It is based not on comparison of the repeated sound against a pre-existing database, such as hidden Markov models (HMM) for speech recognition [24], but only on acoustic comparison of phonemes, and could thus be used universally, independently of the speaker's language.

The present study found a significant negative correlation between manual and automated phoneme scoring, in both normal-hearing ($r = -0.84$) and hearing-impaired subjects ($r = -0.87$), in line with the study hypothesis. Supporting the universality of this automated scoring, Vaerenberg et al. [8] reported similar correlations in Flemish ($r = -0.84$), Dutch ($r = -0.79$), German ($r = -0.85$) and Portuguese ($r = -0.70$) populations, with DTW scores ranging from 29 to 58 (compared to 24 to 53 in the present study). DTW scores were better in the present study than in other populations tested, very probably due to technical testing conditions, with a soundproof booth in the present study, whereas Vaerenberg used a quiet environment but not an actual booth. It is important to take account of test conditions, to ensure reproducibility.

Vaerenberg et al. demonstrated that automated scoring showed the same psychometric properties as classical, manually scored speech audiometry curves: the intelligibility threshold found in 10 normal-hearing subjects was 19.5 dB on automated scoring, and 20 dB on manual scoring of the Bruges CVC logatom test [8].

Objective automated scoring is thus available for self-administered speech audiometry. Customized lists provide results equivalent to those with pre-existing lists, but raise a number of questions.

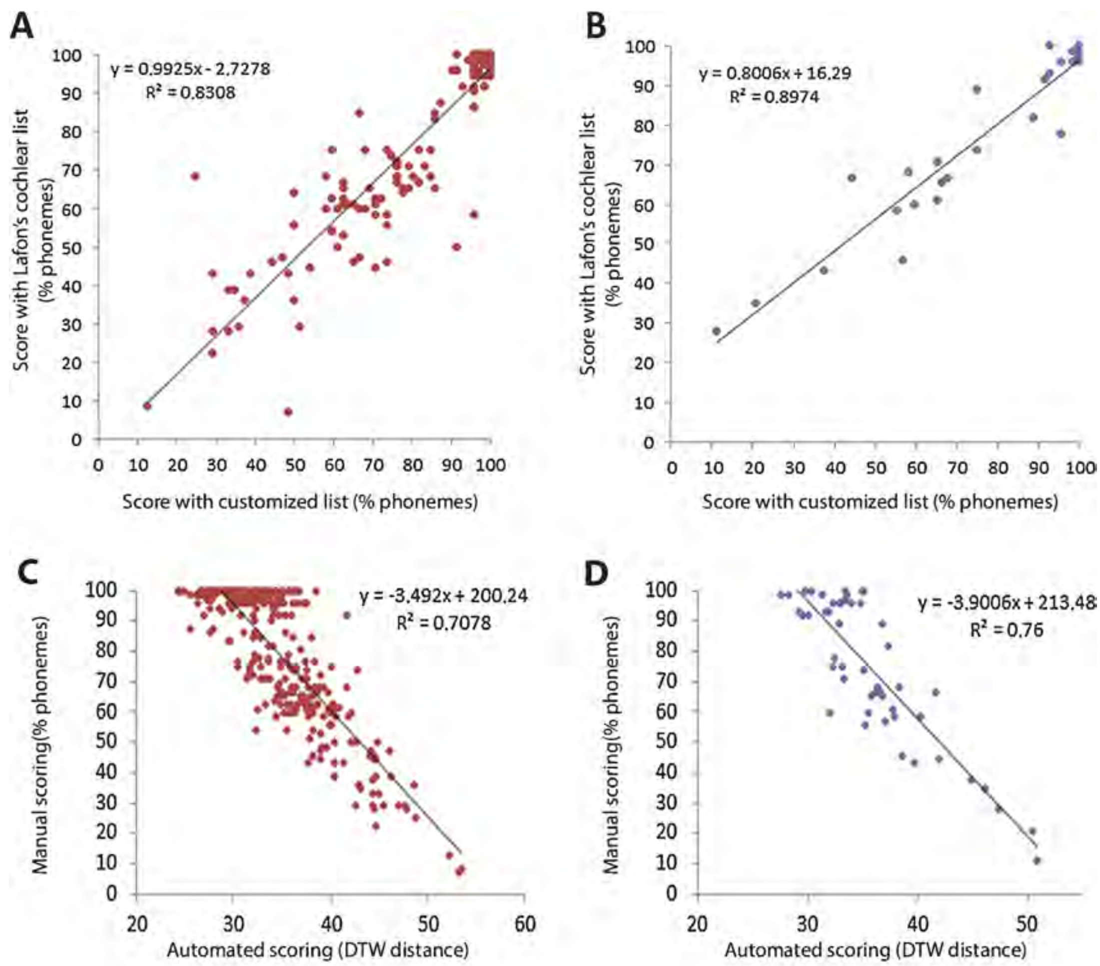

Fig. 3. Correlation between scores with Lafon cochlear lists and self-recorded customized lists and between automated scoring (DTW distance) and manual scoring of customized lists. There was a highly significant positive correlation between customized and Lafon list scores in both populations: (A) normal-hearing ($n = 77$) and (B) hearing-impaired ($n = 13$). Likewise, there was a strong negative correlation between automated and manual scoring of customized lists in (C) the normal-hearing ($n = 77$) and (D) hearing-impaired ($n = 13$) populations.

The time needed to create such lists may make the method impractical for the purposes of a single diagnosis. For clinical or prosthetic audiological monitoring, on the other hand, the time taken to create the lists is usefully made up for by repeated assessment, especially in following up a hearing aid or cochlear implant. Moreover, unlike most lists available in French, the OTOSPEECH package works from texts familiar to the individual subject, taking account of the occurrence of phonemes in the subject's language [25], thus ensuring that list phonemes are representative, all of the words being familiar to the subject. This should limit the inter-list and inter-subject variation found notably with 2-syllable lists such as those of Fournier, and make the test closer to everyday life conditions.

Although results seem promising, the present study compared just 2 customized and 2 prerecorded lists. Larger-scale validation on a larger number of customized lists will be needed before this approach can be recommended in French. The software should also be validated for speech audiometry in noise.

5. Conclusions

Speech audiometry using customized lists, self-voice recording and automated scoring gave results equivalent to those of traditional speech audiometry using prerecorded lists and manual scoring.

This approach could be implemented in audiological monitoring of hearing-aid or cochlear implant bearers, avoiding memory bias and reducing variability of measurement.

Disclosure of interest

The Otospeech software package is produced by Otoconsult NV. BV is employed by Otoconsult and PG receives intellectual property royalties on Otoconsult products.

Acknowledgments

The authors thank Dr Thibault Mura (Medical Information Dept., Montpellier University Hospital) for re-editing and for validation of the statistical tests.

References

[1] Hirsh IJ, Davis H, Silverman SR, Reynolds EG, Eldert E, Benson RW. Development of materials for speech audiometry. J Speech Hear Disord 1952;17(3):321–37.
[2] Gelfand JT, Christie RE, Gelfand SA. Large-corpus phoneme and word recognition and the generality of lexical context in CVC word perception. J Speech Lang Hear Res 2014;57(1):297–307.
[3] Werker JF, Tees RC. Phonemic and phonetic factors in adult cross-language speech perception. J Acoust Soc Am 1984;75(6):1866–78.
[4] Hammer A, Vaerenberg B, Kowalczyk W, ten Bosch L, Coene M, Govaerts PJ. Balancing word lists in speech audiometry through large spoken language corpora. In: INTERSPEECH; 2013. p. 3613–6.

[5] Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. Acoustics, Speech and Signal Processing, IEEE Transactions on 1978;26(1):43–9.
[6] Darch J, Milner B, Vaseghi S. Analysis and prediction of acoustic speech features from mel-frequency cepstral coefficients in distributed speech recognition architectures. J Acoust Soc Am 2008;124(6):3989–4000.
[7] Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. Acoustics, Speech and Signal Processing, IEEE Transactions on 1980;28(4):357–66.
[8] Vaerenberg B, ten Bosch L, Kowalczyk W, Coene M, Smet HD, Govaerts PJ. Language-universal speech audiometry with automated scoring. In: INTERSPEECH; 2013. p. 3608–12.
[9] Dodele L, Dodele D. L'audiométrie vocale en présence de bruit et le test AVfB. Cah Audition 2000;13(6):15–22.
[10] Lafon JC. [Phonetic test, phonation, audition]. J Fr Otorhinolaryngol Audiophonol Chir Maxillofac 1972;21(3):223–9.
[11] Jansen S, Koning R, Wouters J, van Wieringen A. Development and validation of the Leuven intelligibility sentence test with male speaker (LIST-m). Int J Audiol 2014;53(1):55–9.
[12] Preece JP, Fowler CG. Relationship of pure-tone averages to speech reception threshold for male and female speakers. J Am Acad Audiol 1992;3(3):221–4.
[13] Plyler PN, Alworth LN, Rossini TP, Mapes KE. Effects of speech signal content and speaker gender on acceptance of noise in listeners with normal hearing. Int J Audiol 2011;50(4):243–8.
[14] Bishop J, Keating P. Perception of pitch location within a speaker's range: fundamental frequency, voice quality and speaker sex. J Acoust Soc Am 2012;132(2):1100–12.
[15] Kilic MA, Ogut F. The effect of the speaker gender on speech intelligibility in normal-hearing subjects with simulated high frequency hearing loss. Rev Laryngol Otol Rhinol (Bord) 2004;125(1):35–8.
[16] Johnson K. Differential effects of speaker and vowel variability on fricative perception. Lang Speech 1991;34(Pt 3):265–79.
[17] Zaske R, Schweinberger SR. You are only as old as you sound: auditory after effects in vocal age perception. Hear Res 2011;282(1–2):283–8.
[18] Hay J, Drager K, Warren P. Short-term exposure to one dialect affects processing of another. Lang Speech 2010;53(Pt 4):447–71.
[19] Hanulikova A, van Alphen PM, van Goch MM, Weber A. When one person's mistake is another's standard usage: the effect of foreign accent on syntactic processing. J Cogn Neurosci 2012;24(4):878–87.
[20] Strombergsson S. Children's recognition of their own recorded voice: influence of age and phonological impairment. Clin Linguist Phon 2013;27(1): 33–45.
[21] Candini M, Zamagni E, Nuzzo A, Ruotolo F, Iachini T, Frassinetti F. Who is speaking? Implicit and explicit self and other voice recognition. Brain Cogn 2014;92C:112–7.
[22] Aruffo C, Shore DI. Can you McGurk yourself? Self-face and self-voice in audiovisual speech. Psychon Bull Rev 2012;19(1):66–72.
[23] Xu M, Homae F, Hashimoto R, Hagiwara H. Acoustic cues for the recognition of self-voice and other-voice. Front Psychol 2013;4:735.
[24] Rose RC, Juang BH. Hidden Markov models for speech and signal recognition. Electroencephalogr Clin Neurophysiol Suppl 1996;45:137–52.
[25] Coene M, Hammer A, Kowalczyk W, Ten Bosch L, Vaerenberg B, Govaerts PJ. Quantifying cross-linguistic variation in grapheme-to-phoneme mapping. INTERSPEECH 2013:1854–7.