

Developing an Arabic speech intelligibility test for adolescents and adults

Mona A. Hegazi^a, Ahmed Abdelhamid^{b,c,d,e}

^aPhoniatrics Unit of Phoniatrics, Otorhinolaryngology Department, Faculty of Medicine, ^bAssistant Professor of Phoniatrics, College of Medicine, Ain Shams Hospitals, Cairo, Egypt, ^cConsultant of Phoniatrics, ENT Department, King Fahad Hospital of the University, AlKhobar, ^dAssistant Professor of Phoniatrics, ^eENT Department, College of Medicine, Imam Abdulrahman Bin Faisal University (IAU), Dammam, Saudi Arabia

Correspondence to Ahmed Abd El-Hamid, MD, Ph.D, 18 El-Zaitoun Station Street, Cairo, 31952, Egypt, postal code 40148. Tel: +20 100 771 1106/+20 966 568 360 191; fax +966133330333; e-mail: dr_mido1976@yahoo.com

Received 6 January 2018

Accepted 6 June 2018

The Egyptian Journal of Otolaryngology
2019, 35:86–91

Objectives

The improvement of speech intelligibility of many patients is one of the primary aims of the therapy of communication disorders. The standard evaluations lack an Arabic test to measure speech intelligibility among adolescents and adults.

Participants and methods

This study was conducted on 200 participants with an age range from 12 to 60 years who can read Arabic. All participants were randomly selected from the outpatient clinic of phoniatrics from five speech disorders affecting speech intelligibility. Each participant included in the study was subjected to two evaluations: a subjective rating of the participant's speech intelligibility and the developed Arabic speech intelligibility test, which is meant to be an objective measure.

Results

The results showed highly significant correlation between the scores of the Arabic speech intelligibility test and the average scores of the raters.

Conclusion

The developed test proved to be valid and reliable for measuring speech intelligibility and could be categorically classified into ranges of severity.

Keywords:

Arabic test, communicative disorders, objective, rating, reliability, speech intelligibility, validity

Egypt J Otolaryngol 35:86–91

© 2019 The Egyptian Journal of Otolaryngology

1012-5574

Introduction

Speech intelligibility is a measure of the understandability of a speech sample. It is an important measure of the functional limitations experienced by speakers with communication disorders [1]. The improvement of intelligibility of many patients is one of the primary aims of the therapy of communication disorders. Examples of such disorders that reduce speech intelligibility are hearing impairment (HI), dysarthria, speech sound disorders, velopharyngeal dysfunction (VPD), alaryngeal speech, and some cases of fluency disorders (FDs). Assessment of the speech intelligibility is also one means of monitoring the efficacy of therapy of these patients and to assess the efficiency of some devices such as hearing aids.

Until recently, patients' speech intelligibility has been assessed depending on subjective methods by commenting on the overall speech during patient's interviewing or during recitation of memorized verses whether directly or from audio-tapes. The results of the subjective methods are affected by many factors. For example, the listener's experience with a certain speech disorder increases his or her evaluation of the patient's intelligibility. The familiarity of the listener with the material spoken and the context in which it is said increases the

subjective impression of the patient's intelligibility. In addition, the familiarity of the listener with the speaker and his communication may affect judgment of intelligibility. Sometimes the visibility of the speaker to the listener increases the listener's ability of understanding him/her. This makes judging intelligibility from audio-tapes less granted than that done while facing the patient. Monsen [2] discussed the magnitude of the previous factors as well as other factors on judging of speech intelligibility. Other factors on the speaker's side – such as the complexity of the spoken message, the rate of speech, and whether the spoken material is practiced or generated spontaneously – all affect a listener's judgment of intelligibility. Therefore, stabilization of all factors as well as standardization of a speech material is important to facilitate judgment of intelligibility and to allow better comparisons across therapy.

To date, there are no Arabic tests that measure speech intelligibility in adolescents and adults. Thus, it was necessary to develop and standardize an objective Arabic speech intelligibility test that can be used to

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

estimate and score the degree of speech intelligibility in communication disorders in these age categories.

The aim of this work is to construct a valid and reliable Arabic speech intelligibility test that can be used in evaluating the efficacy of different therapy programs in the different communication disorders in which speech intelligibility is affected.

Participants and methods

Participants

This study was conducted on 200 patients with an age range from 12 to 60 years. They were 123 males and 77 females. They were randomly selected from the outpatient clinic of phoniatrics from the following disorders: motor speech disorders, VPD, FDs, multiple speech sound production disorders, and HI participants. Forty patients from each disorder were included.

Inclusion criteria

The following were the inclusion criteria:

- (1) Literate.
- (2) Average mentality.
- (3) Acceptable language skills.

Exclusion criteria

The following were the exclusion criteria:

- (1) Illiterate participants.
- (2) Marked visual problem.
- (3) Subnormal mentality.
- (4) Poor language skills.

Methods

Each participant included in the study was subjected to two evaluations:

- (1) A subjective rating of the participant's speech intelligibility: this was done by three naive persons who were not familiar with the patient (R1, R2 and R3). To avoid the element of past experience with the participant's speech and the nature of the disorder, these raters were neither phoniatrists nor logopedists. After establishing rapport with the clinician, the participant was asked to talk about a certain topic in ~3–5 min. Then the three outside raters were admitted to the room and allowed to listen to the subject without prior knowledge about the topic that he/she was talking about. Then all three raters were asked to rate for the intelligibility along a five-point scale (Appendix A) [3].

- (2) Test application: the test is composed of 250 cards carrying 125 words (each word is repeated twice). The words are structurally organized into five sets. Each set consists of 25 phonetically confusable words which are all real words. Phonetically confusable are like minimal pairs (as seen in Table 3, e.g. fas, bas, and mas). They may differ only in one consonant or one vowel. This adds challenge to the listener to try to discriminate correctly what is being uttered by the patient. They are segregated as follows:

- (a) Set A includes 25 (×2) red cards of monosyllabic words which start with: bilabial and labiodental consonants (/m/, /b/, /f/).
- (b) Set B includes 25 (×2) green cards of monosyllabic words which start with: interdental, linguo-dental and linguo-alveolar consonants (except/s/) (/t/, /d/, /t̪/, /d̪/, /l/, /n/, /θ/).
- (c) Set C includes 25 (×2) yellow cards of monosyllabic words which start with: /s/, post-alveolar and palatal consonants (/s/, /s̪/, /ʃ/, /z/, /r/).
- (d) Set D includes 25 (×2) white cards of monosyllabic words which start with: linguo-velar and linguo-uvular consonants (/k/, /g/, /x/, /ŋ/, /θ/).
- (e) Set E includes 25 (×2) blue cards of monosyllabic words which start with: pharyngeal and glottal consonants (/ħ/, /ʕ/, /h/, /ʔ/).

The test is constructed following the general guidelines proposed by Monsen [4] and Chin *et al.* [5]. The examiner is seated facing the patient in a quiet room with the cards placed on a table, face down, between the participant and the examiner. The cards are scrambled before starting. Then the participant will pick each word and he/she reads it, without the examiner seeing the word or looking at the participant being tested. Each set is tried separately. Then the examiner will write down – in order – what he/she thinks was uttered by the participant in a scoring sheet. Then the cards are placed – in the same order – in a separate box to be matched with the examiner's sheet for estimation of the number of correct responses. The number of correct responses in relation to the total number of words (250) is expressed in percentage.

The Arabic speech intelligibility test for adolescents and adults (ASIT-AA) is designed to provide an estimation of the overall speech intelligibility by providing a total score in percentage by relating the number of correct responses to the total number of responses (250).

Thirty participants were chosen randomly and retested with the same test after an interval of 3 weeks.

Data management and analysis

The statistical package for the social sciences [6] was used for analysis of data. Descriptive statistics were shown as mean, \pm SD, percentages, and ranges. Cohen's κ coefficient was used to measure inter-rater agreement for categorical items. 'Substantial agreement' is considered when κ is between 0.61 and 0.80 and 'almost perfect agreement' when κ is between 0.81 and 1.00 [7]. Cronbach's α test was used to detect test-retest agreement for continuous variables. Pearson correlation was used to measure the strength and direction of the relationship between pairs of continuous variables with r value denoting the strength and the direction of the relationship. Spearman's rank order correlation was used to assess the strength of association between two variables, one of them is ordinal. The correlation coefficient, denoted symbolically as ' r_s ', defines the strength and direction of the relationships.

Results

Demographic data showed that patients' mean age was 25.08 ± 12.61 years with a range from 12 to 60 years. According to sex, 123 (61.5% of patients) were males and 77 (38.5%) were females.

Agreement among the raters

An inter-rater reliability analysis using the κ statistic was performed to determine consistency among raters. The inter-rater reliability for the raters was found to score κ of a range of 0.87–0.94 ($P < 0.001$). The scores of all three raters showed significant agreement (an 'almost perfect agreement') (Table 1). An average value of the three raters was taken to represent the score of the subjective test.

Correlation between the subjective scores and the Arabic test scores according to disorders

By applying Spearman's rank order correlation, there was a strong positive correlation between the subjective rating of the judges and the test scores as shown in Table 2. The values of each test are demonstrated in Fig. 1a and b.

Table 1 κ Scores testing agreement among raters

Raters	R1	R2	R3
R1	–	0.94	0.85
R2	0.94	–	0.91
R3	0.85	0.91	–

The mean κ across pairs is 0.90.

Test reliability

This was measured in two ways from a sample of 30 participants.

The first way was by measuring the agreement between each word when rated twice in the same testing using κ test. Table 3 shows that the agreement ranged between 0.73 and 1.00 for κ . These results are considered 'substantial agreement' in 17% of words and 'almost perfect agreement' in 83% of the words.

The second way was by the test retest method for the total scores obtained from 30 patients after 3 weeks interval using Cronbach's α test. This revealed a significant ($P < 0.01$) correlation for consistency with an r_s value of 0.87.

The results of κ test for agreement for repeated words within the test from 30 patients indicated a significant agreement ($P < 0.01$) for all items of the test signifying reliability of the test.

Test validity

The total score of the Arabic adults speech intelligibility test versus the average score of the three raters (external validity) and groups of the test (construct validity) were tested for correlation.

As shown in Table 4, it was clear that the correlation between the total score of the Arabic intelligibility test and the subjective score (average score of raters) was highly significant. In addition, the correlation between the total score of the Arabic intelligibility test and its groups was highly significant.

Categorical classification of the scores

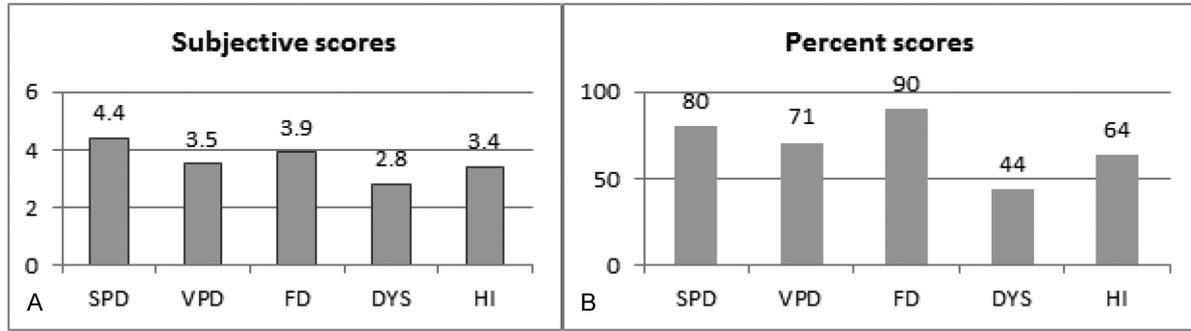
A highly significant correlation between the scores of ASIT-AA and the subjective test's scores was obtained. This fact allowed the final intelligibility scores of the

Table 2 The correlation between the mean score of the Arabic adults' speech intelligibility test and the mean score of subjective test in all participants and participants in each disorder

Disorder	Mean score of raters (the subjective test)	Mean score of Arabic speech intelligibility test	r_s
All participants	3.67	65.62	0.94**
SPD	4.35	80.28	0.89**
VPD	3.46	71.24	0.91**
FD	3.85	90.41	0.84**
DYS	2.78	44.18	0.93**
HI	3.36	64.4	0.92**

DYS, dysarthria; FD, fluency disorders; SPD, sound production disorders; VPD, velopharyngeal dysfunction. ** $P < 0.001$, correlation is significant.

Figure 1



The results obtained in each of the disorders by the subjective testing (a) and by the speech intelligibility test (b). DYS, dysarthria; FD, fluency disorders; HI, hearing impairment; SPD, sound production disorders; VPD, velopharyngeal dysfunction.

Table 3 The results of κ test to determine agreement of the scores obtained in each word when repeated

No.	Set A Test items	K	Set B Test items	K	Set C Test items	K	Set D Test items	K	Set E Test items	K
1.	فأل/fə:l/	0.80	داب/də:b/	0.84	ساب/sə:b/	0.90	قطر/qə:tr/	0.91	عمر/ʔə:mr/	1.00
2.	فأس/fə:s/	0.87	تاب/tə:b/	0.80	شأب/ʃə:b/	0.85	جال/gə:l/	0.82	أذن/ʔə:dn/	0.89
3.	مأس/mə:s/	0.89	نأب/nə:b/	0.93	شأل/ʃə:l/	0.88	غصن/ʔə:sn/	0.91	همس/hə:ms/	0.87
4.	موز/mo:z/	0.97	لام/lə:m/	0.83	سأل/sə:l/	0.75	غار/ʔə:r/	0.94	عوم/ʔə:m/	0.95
5.	بأس/bə:s/	0.93	نأم/nə:m/	0.85	صار/sə:r/	0.84	كلب/kəlb/	0.95	هدم/hədm/	0.83
6.	فأر/fə:r/	0.99	دال/də:l/	0.93	رأب/rə:b/	0.86	جار/gə:r/	0.88	عام/ʔə:m/	0.89
7.	بر/bə:r/	1.00	ذال/də:l/	0.87	صمغ/sə:mʔ/	0.80	خاف/xə:f/	0.89	هام/hə:m/	0.90
8.	فر/fə:r/	0.86	نأل/nə:l/	0.93	صمت/sə:mt/	0.79	كوز/ku:z/	0.90	حار/ħə:r/	0.87
9.	مر/mə:r/	0.90	ثوم/to:m/	0.80	صبر/sə:br/	0.86	كأس/kə:s/	0.90	عين/ʔə:n/	0.97
10.	مصر/mə:sr/	0.95	طأر/tə:r/	0.93	شعر/ʃə:ʔr/	0.96	كيس/ki:s/	0.94	عأر/ʔə:r/	0.86
11.	فصل/fə:sl/	0.84	نأر/nə:r/	0.89	سيف/se:f/	0.94	غين/ʔe:n/	0.80	أمر/ʔə:mr/	0.92
12.	بحر/bə:ħr/	0.91	دأر/də:r/	0.79	سيل/se:l/	0.84	قلب/qəlb/	0.89	أمس/ʔə:ms/	0.90
13.	فول/fu:l/	0.89	طأف/tə:f/	0.82	شيل/ʃi:l/	0.86	قبر/qə:br/	0.88	هجر/hə:ʔr/	0.89
14.	ببر/bi:r/	0.87	دبر/de:r/	0.90	سبب/si:b/	0.87	كان/kə:n/	0.87	أحن/ħə:n/	0.89
15.	فأل/fə:l/	0.91	نوم/no:m/	0.85	ربع/robə:ʔ/	0.91	جأف/gə:f/	0.79	أمن/ʔmn/	0.93
16.	فيل/fi:l/	0.78	تأمر/tə:mr/	0.91	شيل/ʃi:l/	0.88	قطن/qə:tn/	0.89	أهل/ʔəhl/	0.88
17.	بأرد/bə:rd/	0.90	تأل/ti:l/	0.84	سور/su:r/	0.90	قأصر/qə:sr/	0.90	عأصر/ʔə:sr/	0.90
18.	وأرد/wə:rd/	1.00	نأل/ni:l/	0.84	رأب/roə:b/	1.00	خأبط/xə:bt/	0.92	أصل/ʔə:sl/	1.00
19.	وأحل/wə:ħl/	0.87	تأين/ti:n/	0.80	رأب/roə:b/	0.90	عأب/ʔə:b/	0.91	عأبد/ʔə:bd/	0.87
20.	وعأل/wə:ħl/	0.89	دأك/di:k/	0.93	زأغ/zə:ʔ/	0.89	قأهر/qə:ħr/	0.82	أكل/ʔə:kl/	0.89
21.	بأور/bu:r/	0.75	تأور/to:r/	0.73	صأب/sə:bb/	0.80	خأب/xə:b/	0.88	هأن/hə:n/	0.85
22.	مأهر/mə:ħr/	0.93	نأول/no:l/	0.85	سأب/sə:bb/	0.91	خأبز/xə:bz/	0.80	حأرب/ħə:rb/	0.93
23.	وأصل/wə:sl/	0.79	ضأف/de:f/	0.93	سأر/sə:r/	0.90	غأبط/ʔe:bt/	0.84	أأطر/ʔə:tr/	0.96
24.	مأوت/mo:t/	0.90	طأف/te:f/	0.87	زأيت/ze:t/	1.00	خأل/xə:l/	0.85	حأوض/ħə:wd/	1.00
25.	فأجر/fə:ʔr/	0.88	نأور/nu:r/	0.93	ظأرف/zə:rf/	0.86	كأف/kə:f/	0.93	حأصر/ħə:sr/	0.80

All κ values are significant.

Table 4 The correlation between total scores of the Arabic speech intelligibility test for adolescents and adults versus average scores of raters and the test subgroups

	Mean of the three raters' scores	Set A	Set B	Set C	Set D	Set E
Total score of ASIT-AA mean value=65.62						
Mean value	3.67	36.69	37.43	31.99	34.55	38.78
r Value	0.94	0.95	0.95	0.81	0.87	0.86
Significance	$P<0.01$	$P<0.01$	$P<0.01$	$P<0.01$	$P<0.01$	$P<0.01$

$P<0.01$, highly significant.

Table 5 The categorical classification of speech intelligibility according to the scores of Arabic adults' speech intelligibility test

Score of the subjective test	%Range of the Arabic speech intelligibility test scores	Intelligibility
1	0–24	Unintelligible speech
2	25–59	Poor intelligibility
3	60–74	Fair intelligibility
4	75–89	Good intelligibility
5	90–100	Excellent intelligibility

test to be expressed in a categorical classification as shown in Table 5. The categories are unintelligible speech, poor intelligibility, fair intelligibility, good intelligibility, and excellent intelligibility. The classification depended on Z-scores and 90% confidence intervals obtained in correspondence to each category of the subjective test.

Discussion

Clinicians usually ask for more objective ways to quantify conversational intelligibility in communication disorders. This is helpful to assess the severity of the disorders and to evaluate efficacy of therapy programs. In general, two approaches have been used in evaluating speech intelligibility. The first is a 'listener rating' approach for a list of words, or sentences read by the patient on a rating scale. This approach has the disadvantages of often having poor reliability among raters and their ratings are a mere reflection of their experience with the type of the patient's speech. The second approach is the 'listener's response' or 'item identification' approach where listeners are required to record all the intelligible words in variable sets of spoken sentences. Although this approach is more objective, but it is time consuming [8,9].

In this study, participants having one of five clinical communication disorders were used to standardize an Arabic test for measuring speech intelligibility. The reason for diminished intelligibility is different in each disorder. Causes for the reduced intelligibility include affection of the speaker's vocal intensity, frequency, quality, speech rate, inflections, pauses, stress patterns, fluency, articulation, resonance, and prosody. Some of these factors may contribute in reduced intelligibility with variable degrees in each disorder. Among the five clinical disorders studied, patients with FDs scored the highest intelligibility scores when assessed by the ASIT-AA, followed by sound production errors,

then VPD, then HI, and lastly dysarthria (Fig. 1). Using the subjective scores, FDs and articulatory errors switched places in the ranking. This switch may be due to the difference in the test procedure; people who stutter perform better in tasks requiring short responses as well as in reading tasks. On the contrary, the highest intelligibility scores were obtained in sound production errors group on the subjective ratings because listeners usually quickly accustom their brain to the errors they perceive in the speaker's conversation. Hence, the higher scores in contextual ratings. The breakdown in FDs is more manifest in longer sentences and in conversation owing to the burden they pose to the speaker. This explains why the lowest agreement results between both methods of testing were obtained in FDs.

The authors believe that the gold standard against which a test for speech intelligibility can depend on for testing validity is when judged by naive listeners. This situation is similar to the real-world communicative situations which speakers face. Understanding any language by the listener is accomplished greatly by the context of the conversation, of course in addition to knowledge of the language. For these reasons, the subjective rating representing sentence-level open-set assessment was performed by judges who do not know the participant, nor are they personnel who deal with these patients in any way and who may be aware of the speech patterns of the patients they serve. Topics that the patients were asked to speak about were changed every time the same judge was told to rate any patient's speech. The scale used here in that subjective rating was meant to be a narrow scale of only five grades, with clear boundaries between each grade. In this way, the disadvantage of inaccuracy in using a wide scale is avoided [10], giving the test a quasi-objective nature rather than being purely subjective. Relating the results of the newly designed ASIT-AA to this scale thus renders the test concurrent validity (criterion-related external validity). Construct validity was proved by the significant correlation between the five sets of the test and the total score.

The test is designed so that all words make sense. The words were chosen to be monosyllabic words to make distinction targeted at one aim as much as possible. It is known that sentence contexts greatly enhance the accuracy of word recognition relative to isolated words [11]. Similarly, words with complex morphology may provide listeners with more contextual cues for phoneme identification than do simpler, monosyllabic words. Thus,

monosyllabic words offer the least opportunity for listeners to use lexical knowledge to aid word recognition [12].

In each set, the words used aimed at testing contrasts between voiced and nonvoiced consonants (e.g. /tθ:b/, /dθ:b/), manner contrasts (e.g. /mθs/, /bθs/), and consonant position contrasts (e.g. /fA:r/, /bA:r/). Word contrasts with vowels differing in tongue height (e.g. /bu:r/, /bA:r/, /bi:r/) and vowel length (e.g. /bArr/, /bA:r/) were used. Vowels' distinction is considered one of the factors that improve the speech intelligibility of the HI and dysarthric speakers [13]. In open nasality, the vowel height affects the degree of open nasality of such patients [14]. These variations included in the test design render it content validity. Test words were repeated twice in the design to minimize the effect of guessing by exclusion on the examiner's side. The examiner's role was to write down what he heard from the participant without relying on additional visual cues. His concern should not be how well the word was said but only whether the quality of articulation was functionally good enough to make the word understandable. The time of administration of the test ranged from 40 to 45 min. Test reliability was tested two ways. The first is for each word in the test when repeated twice on the same list. The second way is by evaluating the agreement of the total scores when performed on the same participant 3 weeks later (test-retest reliability). The high agreements obtained in the two procedures prove the consistency and stability of the test and hence high reliability.

The test is thus designed on the basis of determining the percentage of correctly understood words by the examiner from a closed set that carries monosyllabic word contrasts without visualizing the patient being evaluated. In this way, the number of variables that may affect speech intelligibility are minimized. These variables are proved to affect listeners' responses [15–17].

Conclusion

This work presents a valid and reliable Arabic adult speech intelligibility test that can be used in assessment and follow-up of communication disorders. It can be expressed either in percentage scores or in categorical scores based on the obtained percentages.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

Appendix

Table A1 Subjective impression of the patients' speech intelligibility

Grade	What the grade stands for
5	I can understand all patient's speech (normal intelligibility)
4	I can understand most of the patient's speech but I miss occasional words. As a whole I can understand what he intends to say (good intelligibility)
3	I have great difficulty in understanding what the patient is saying. As a whole I may understand what he is talking about (fair intelligibility)
2	I find extreme difficulty in understanding what the patient is talking about at all but I can pick some words which I can understand (poor intelligibility)
1	I cannot understand anything at all (unintelligible speech)

References

- 1 Yorkston K, Strand E, Hume J. The relationship between motor function and speech function in amyotrophic lateral sclerosis. In: Cannito M, Yorkston KM, Beukelman DR, editors. *Neuromotor speech disorders: nature, assessment, and management*. Baltimore: Paul H. Brookes 1998. pp. 85–98.
- 2 Monsen RB. The oral speech intelligibility of hearing – impaired talkers. *J Speech Hear Res* 1983; 48:286–296.
- 3 Flipsen P Jr. Measuring the intelligibility of conversational speech in children. *Clin Linguist Phon* 2006; 20:202–312.
- 4 Monsen RB. A usable test for the speech intelligibility of deaf talkers. *Am Ann Deaf* 1981; 126:845–852.
- 5 Chin S, Finnegan K, Chung B. Relationships among types of speech intelligibility in pediatric users of cochlear implants. *J Commun Disord* 2001; 34:187–205.
- 6 SPSS 15.0 Command Syntax Reference. Chicago, Illinois: SPSS Inc. 2006.
- 7 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159–174.
- 8 Monsen R, Moog JS, Geers AE. CID picture spine. *Speech intelligibility evaluation*. St Louis, MO: Central Institute for the Deaf; 1988.
- 9 Kent RD, Weismer G, Kent JF, Rosenbek JC. Toward phonetic intelligibility testing in dysarthria. *J Speech Hear Res* 1989; 54:482–499.
- 10 Schiavetti N. Scaling procedures for the measurement of speech intelligibility. In: Kent RD, editor. *Intelligibility in speech disorders*. Philadelphia, PA: John Benjamins 1992. pp. 11–34.
- 11 Boothroyd A, Nittrouer S. Mathematical treatment of context effects in phoneme and word recognition. *J Acoust Soc Am* 1988; 84:101–114.
- 12 Francis A, Nusbaum H. The effect of lexical complexity on intelligibility. *Int J Speech Technol* 1999; 3:15–25.
- 13 Kewley-Port D, Burkle TZ, Lee JH. Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *J Acoust Soc Am* 2007; 122:2365–2375.
- 14 Beddor PS. (1983). *Phonological effects of nasalization on vowel height: universal patterns*. Unpublished doctoral dissertation. University of Minnesota. Bloomington: Indian University Linguistics Club.
- 15 Anderson Gosselin P, Gagné JP. Older adults expend more listening effort than young adults recognizing speech in noise. *J Speech Lang Hear Res* 2011; 54:944–958.
- 16 Picou EM, Ricketts TA, Hornsby BWY. Visual cues and listening effort: individual variability. *J Speech Lang Hear Res* 2011; 54:1416–1430.
- 17 Nagle KF, Eadie TL. Listener effort for highly intelligible tracheoesophageal speech. *J Commun Disord* 2012; 45:235–245.