

# Medidas de inteligibilidade nos distúrbios da fala: revisão crítica da literatura\*\*\*

## Intelligibility measurements in speech disorders: a critical review of the literature

Simone dos Santos Barreto\*  
Karin Zazo Ortiz\*\*

\*Fonoaudióloga. Mestre em Distúrbios da Comunicação Humana pela Universidade Federal de São Paulo. Fonoaudióloga da Prefeitura do Rio de Janeiro. Endereço para correspondência: Rua Botucatu, 802 - São Paulo - SP - CEP 04023-062 (simone\_barret@hotmail.com).

\*\*Fonoaudióloga. Pós-Doutorado em Neurociências pela Universidade Federal de São Paulo. Professor Adjunto do Departamento de Fonoaudiologia da Universidade Federal de São Paulo.

\*\*\*Trabalho Realizado no Departamento de Fonoaudiologia da Universidade Federal de São Paulo - Escola Paulista de Medicina.

Artigo de Revisão de Literatura

Artigo Submetido a Avaliação por Pares

Conflito de Interesse: não

Recebido em 26.11.2007.  
Revisado em 20.03.2008.  
Aceito para Publicação em 10.06.2008.

### Abstract

Background: the reduction in speech intelligibility is considered one of the main characteristics of individuals with speech disorders, and is an important issue for clinical and research investigation. In spite of its relevance, the literature does not present a consensus on how to measure speech intelligibility. Besides the diversity of existent methods, another important issue refers to the influence of certain variables on these measurements and, consequently, on the interpretation of the results. Aim: to investigate evidence on the agreement between speech intelligibility measurements, obtained through different methods, used in the assessment of speech disorders, and to identify the effect of variables related to assessment procedures or to the listener. A critical review of articles indexed in the databases Medline, Web of Science, Lilacs and Scielo, until October 2007, was carried out. The key-word used to perform the search was speech intelligibility. Conclusion: there was no evidence of agreement between the speech intelligibility measurements obtained through different methods in the investigated literature. This fact limits the comparison between clinic and research results on speech intelligibility of individuals with speech disorders. Besides that, it was observed that some variables can interfere in these measurements, such as: type of task and speech stimulus, signal presentation mode, type of required answer and listener's experience with the speaker. These must be considered when interpreting the results of speech intelligibility tests.

**Key Words:** Speech Intelligibility; Speech Production Measurement; Speech Disorders; Speech.

### Resumo

Tema: a redução da inteligibilidade da fala é considerada uma das principais manifestações encontrada em sujeitos com distúrbios da fala, sendo um importante objeto de investigação fonoaudiológica. Apesar de sua relevância, não existe consenso na literatura da área de como a inteligibilidade da fala deva ser avaliada. Além da questão da diversidade de métodos existentes, outro aspecto importante refere-se à influência que determinadas variáveis podem exercer sobre tais medidas e, conseqüentemente, sobre sua interpretação. Objetivo: investigar a existência de possíveis evidências acerca da concordância entre medidas de inteligibilidade obtidas por diferentes métodos de mensuração, empregados na avaliação de sujeitos com distúrbios da fala, e identificar os efeitos de variáveis relacionadas aos procedimentos de avaliação ou ao ouvinte sobre essas medidas. Para tal, foi realizada uma revisão crítica de artigos sobre o tema, indexados nas bases de dados Medline, Web of Science, Lilacs e Scielo, até outubro de 2007, através dos termos de busca speech intelligibility ou inteligibilidade da fala. Conclusão: não foram encontradas evidências, na literatura pesquisada, de concordância entre as medidas de inteligibilidade da fala obtidas por métodos distintos, o que limita a comparação entre resultados clínicos e de pesquisas sobre inteligibilidade em sujeitos com distúrbios da fala. Além disso, constatou-se que algumas variáveis podem interferir nessas medidas, como: a tarefa e o estímulo de fala, seu modo de apresentação, o tipo de resposta requerido e a experiência do ouvinte com o falante, as quais devem ser consideradas na interpretação dos resultados dos testes de inteligibilidade.

**Palavras-Chave:** Inteligibilidade da Fala; Medidas de Produção da Fala; Distúrbios da Fala; Fala.

Referenciar este material como:



Barreto SS, Ortiz KZ. Intelligibility measurements in speech disorders: a critical review of the literature (original title: Medidas de inteligibilidade nos distúrbios da fala: revisão crítica da literatura). Pró-Fono Revista de Atualização Científica. 2008 jul-set;20(3):201-6.

## Introduction

Intelligibility of speech can be defined as the degree to which the message of a speaker can be decoded by the listener(1). In other words, it refers to the ease in which a listener is able to comprehend the speech of their interlocutor. Accordingly, intelligibility should not merely be considered an attribute of speech but dependent on a range of listener-associated variables (2) and the context in which the communication takes place(1,3).

Considered one of the main manifestations in subjects with acquired or developmental speech disturbances, low speech intelligibility is an important subject of investigation and intervention in speech therapy(1,4). However, a range of different methods tend to be employed to measure intelligibility of speech, with no consensus having been reached in the literature on assessment technique(1).

Besides the issue of the diversity of methods available, another important aspect is the influence certain variables can have on these measures(1,2). A number of studies on the intelligibility of speech have addressed the effects of several of these variables including: the task employed for speech sample collection(5), the stimulus type used(3,4,6-14), the mode in which samples are presented(11,13,15-17), the response type required to identify the stimuli(7,18), the type of transcription analysis(2), the gender of the listener or rater (13,19), or their familiarity with the speaker(s)(9,11,13,20,21).

Given that different methods and their possible confounding variables may produce different intelligibility scores, the aim of the present study was to investigate possible evidence in the literature of agreement among intelligibility measures obtained by different measurement methods used for assessing subjects with speech disturbances. Additionally, the study also sought to identify the effects of variables related to the assessment procedure or to the listener, which may interfere in obtaining these measurements.

In pursuit of the study objectives, a critical review of the literature was conducted on secondary data. The material used included articles indexed on the Web of Science, Medline, Lilacs and Scielo databases. The search strategies adopted to identify articles were: search for the term speech intelligibility in the key word fields, words in the title and/or abstract, while covering articles published up until October 2007. Studies which investigated the relationship between different assessment methods for intelligibility of speech were selected, as well as those investigating the

effect of variables related to assessment procedures or the listener, on measures of speech intelligibility. The studies found were all International in nature.

### Methods of speech intelligibility measurement and their agreement

The methods used to assess intelligibility of individuals with speech disturbances can be divided into two groups: scaling methods and item identification methods(1). Scaling methods encompass: direct magnitude estimate(12,19,22), interval scaling(7,23-26), rating of speech sample pairs(1), percentage estimates(7,17) and the analogic visual scale (20). Item identification methods involve multiple formats according to the type of response required to identify speech stimuli. Some variations on each of these methods can be found(1-11,13-18,20,21,23,27-29).

Direct magnitude estimation is a grading method based on a standard speech sample pre-selected by the researcher, to which an intelligibility value is attributed. Listeners subsequently establish values for the speech samples of the subjects assessed, which represent the degree of intelligibility of these utterances in relation to the standard sample(22). In one variant of this method, the listener is asked to designate a value to the first sample assessed, which is then taken as a reference for assessing the remaining items(12,19).

Another commonly employed method of grading intelligibility is interval scaling. In this case, the listener attributes a number to each sample which represents a linear increment on a graduated intelligibility scale. Such scales tend to vary in terms of number of levels under which the intelligibility can be graded(7,23-26).

The rating of speech sample pairs and percentage estimates are less-used methods of grading intelligibility. In the former, the listener is asked to compare two pairs of speech samples and to judge which is the most intelligible(1), whereas in percentage estimates, the listener attributes a percentage value which they deem compatible with the proportion of intelligible words in each sample(7,17). The analogic visual scale is another grading method which has also been used recently. This comprises a vertical line of a given length, the ends of which represent the extremes on an intelligibility continuum, whereby the listener indicates the position on the scale which reflects the intelligibility of the speaker(20).

Methods of identifying items can entail orthographic transcription of speech stimuli by the

listener(2-11,13-18,20,21,23,27-29) or the selection of alternatives using a multiple-choice format(1,7,14,23). There are variations on the orthographic approach such as: partial(5,7,18) (for example, only target phonemes of each word) or full transcription of speech samples(2-4,6-11,13-16,18,20,21,23,27-29), and tally of target stimuli only (2,4,5,7,9,18,28) or of all the stimuli(2,3,6,8,10,11,13-16,20,21,23,27,29) in the sample. In the latter case, the weight of each stimulus on the final score may vary. An example is where words which differ in importance in the sentence are assigned different values(11). Intelligibility scores in item identification methods also diverge in terms of the unit of measure, which may be represented by the number (6,15,18) or percentage of stimuli correctly or incorrectly decoded(2-5,7-11,13,14,16,17,20,21,23,27-29).

Few studies were found in the literature searched that had performed comparative analysis among methods of measuring intelligibility of speech. In one such experiment, carried out in 21 speakers with hearing loss, the relationship among measures resulting from three instruments for assessing the intelligibility of speech was analyzed using Spearman's correlation coefficient. One of these instruments was based on interval scaling while the others involved item identification, one being multiple-choice and the other transcription. Strong correlations were found among results of the tests (greater than 0.84), whereby measures by scale interval demonstrated greater correlation with measures by item identification than among each other(23).

In another study performed for this purpose(7), the relationships among intelligibility scores attained by full transcription of speech stimuli, percentage estimates and representations on a graduated scale, of eight dysarthric speakers were investigated using Kendall's correlation coefficient. Strong correlations were found among the measures of the scaling method and the item identification method (greater than 0.72). A study in four dysarthric speakers(17) compared intelligibility scores by transcription and percentage estimates, evidencing higher transcription scores than percentage estimates.

These studies only analyzed the relationship among intelligibility measures based on different measurement methods, but not in terms of agreement. The issue of suitability of statistical methods employed to assess agreement between two clinical measurement methods was discussed by Bland and Altman(30). They criticized the use of correlation coefficients, arguing that strong correlation between the measurements of two different methods does not always imply agreement, given the possibility of high discrepancy between these scores.

Based on the considerations outlined above, the data found in the literature provided no evidence for the existence of agreement among the methods of intelligibility measurement.

#### Effects of variables related to assessment procedures or to listeners on speech intelligibility measures

Several studies were found which directly or indirectly investigated the effects of certain variables on speech intelligibility measures. Some of these variable are listener-related factors(9,11,13,19,20,21) while others are related to the assessment procedures, and may be specific to particular measurement methods(3-17). The results of this study are described below by variable. The first five pertain to the assessment procedures whereas the last two refer to the listener.

#### Speech task types

Irrespective of the intelligibility measurement method, speech samples of subjects assessed are recorded for later analysis by listeners. These tasks may involve reading speech stimuli aloud(1,4,5,8,10,11,13,15,18,20,23-26,29), repeating them(2,3,5,22,27,28), or spontaneous speech(5,14,16,24-26).

The sole study comparing the effects of five tasks on intelligibility scores by transcription in a dysarthric speaker, revealed that only spontaneous speech differed from the other types (reading, repetition, repeated and spontaneous song). In addition, the scores in repetition and reading were very similar, indicating the possibility of using both tasks in intelligibility tests by transcription to give similar results. However, the difference between scores achieved on spontaneous speech versus reading and repetition showed a tendency of these to overestimate the intelligibility of spontaneous speech, so this finding should be interpreted with caution(5).

#### Speech stimuli type

The type of speech stimuli is an aspect which differs greatly in intelligibility assessment which employ stimuli containing different levels of morphosyntactic complexity and semantic predictiveness, such as: single words(1,3,4,6-8,10,14,18,20,23,25,26), separate sentences with or without meaning(3-5,7,9-15,19,20,22,23,27,29) or sentences in narrative form(2,3,8,14,16,23-26). The

effect of employing different speech stimuli on intelligibility measures has been the subject of investigation by a number of researchers and indirectly observed by others(3,4,6-14).

The influence of syntactic-semantic information on intelligibility measures was noted when sentences with and without meaning produced by a normal speaker were compared according to their intelligibility by direct magnitude estimation. Different estimates were obtained, where these proved higher in meaningful sentences(12). The benefit of these syntactic-semantic cues in decoding altered speech was also seen in subjects with speech disturbances (hearing-impaired and dysarthric), when intelligibility by transcription of single sentences was compared with single words. The researchers evidenced that intelligibility scores by transcription of sentences was higher than for single words(9,10,14), with significant magnitude observed only in the best speakers(3,4,7). The type of sentence, in terms of semantic predictiveness, seems to influence intelligibility measurements, since when sentences with high predictiveness were compared to sentences with low predictiveness, the intelligibility scores of speakers rose. A variation of 16 to 30% in the magnitude of difference was observed among studies, which may have been influenced by the presentation mode of stimuli and the speech impairment severity(13), yet not by the listeners' experience with impaired speech(9).

The effect of cohesion of sentences on intelligibility of scores by transcription, when these are presented in narratives was also verified(8,14), as was its interaction with disturbance severity(3). This effect appeared even stronger and more consistent than others, as it influenced the intelligibility not only of speakers with mild dysarthria but also those with moderate and severe dysarthria. The magnitude of the difference among narratives and other types of stimuli ranged from 10 to 30% on average(3).

Regarding the effect of morphosyntactic complexity, this was assessed through comparison of intelligibility by transcription of word lists which differed in terms of their morphological structure. However, for a group of 10 dysarthrics assessed, no differences were seen among intelligibility scores for the three lists(6). Another experiment involving 10 deaf subjects demonstrated that the presence of polysyllables, of consonant clusters and complex syntax reduced intelligibility scores of words in sentences, whereby this effect was greater among less intelligible speakers (in this group, differences of 17% were found, depending on the complexity of the sentences) and among listeners without

experience of speech produced by deaf individuals(11).

#### Speech sample presentation modes

The mode of presentation of speech samples to listeners, with the provision of only auditory information, or auditory and visual information combined, also impacts the intelligibility scores of speech by transcription. Comparison of presentation modes revealed that under combined conditions, intelligibility scores by transcription of sentences was higher among a laryngectomized group, but not among their controls(15), where this pattern repeated for dysarthrics(13,16,17) and deaf subjects(11).

#### Response type in the identification of speech stimuli

Specifically for the item identification method, the response type required for identification of speech stimuli is a factor which stands out in the assessment of intelligibility, whose influence was also verified(7,18). As outlined earlier, intelligibility measures by item identification may be obtained through partial(5,7,18) or full orthographic transcription of speech material(2-4,6-11,13-16,18,20,21,23,27-29) or multiple choice(1,7,14,23).

In a study involving four speakers with cleft palates and/or velopharyngeal insufficiency, the benefit of partial transcription of target phonemes over full transcription of single words was observed in consonant intelligibility scores, although the statistical significance of this difference was not analyzed(18). However, another study carried out with dysarthric speakers, in which intelligibility scores by full transcription, partial transcription (full words in sentences) and multiple choice were compared, a hierarchy in these measures was observed according to difficulty of the required response type. Intelligibility scores of multiple choice format were highest, followed by partial transcription scores and full transcription scores(7).

#### Transcription Analysis type

Concerning intelligibility measures by transcription, the criteria adopted to score each stimulus may differ. The criteria of phonemic correspondence between the orthographic transcription of the listener and the stimulus produced by the speaker is the most frequently used approach(2,3,13,27), although this can be applied in a more flexible manner(5,11,21). In this case, errors in

phonemic correspondence which do not compromise the meaning of the information transferred are not counted, such as omission or inclusion of number morphemes. In a recent study(2) in dysarthrics, three transcription scoring paradigms were analyzed: the exact phonemic correspondence of all the words, exact phonemic correspondence of information words only (content words and modifiers) and the semantic correspondence of information words. Intelligibility scores resulting from the three types of analysis were shown to differ. Nevertheless, the magnitude of differences found was small, suggesting that this was not significant from a clinical standpoint.

#### Listener gender

With regard to the effects of listener gender on intelligibility ratings by direct magnitude estimation, a study with two normal speakers found no significant differences between mean scores of the listener groups(19). The same was verified through a study in a dysarthric speaker, using orthographic transcription of sentences(13).

#### Familiarity of the listener with the speaker

In relation to familiarity of the listener with the speaker, this may be linked to the experience of the listener with a particular speaker or speech of individuals with a certain speech disorder. Concerning the familiarity of the listener with specific speakers, this aspect was studied in a dysarthric speaker(13) by comparing intelligibility scores by transcription of sentences of listeners previously exposed or otherwise, to the speech of this individual. It was observed that familiarity of listener with the speaker had no influence on their intelligibility scores.

In terms of the experience of the listener with impaired speech in general, the findings of

studies remain disparate, indicating higher transcription scores among listeners with experience with deaf speakers(9,11), yet not with tracheoesophageal subjects(20). The magnitude of the difference for deaf speakers was 10% on average, increasing with sentence complexity and condition severity(11). Another study on dysarthric speakers(21), in which the exposure level of listeners to impaired speech was controlled, sentence intelligibility by transcription proved greater in the group of exposed listeners.

#### Conclusion

The data from the present study allow us to state that there is no evidence in the literature studied, of agreement between speech intelligibility measures obtained using different methods. This finding limits comparison of some results of research on intelligibility in populations with speech impairments. In order to employ these measures in clinical speech therapy, the use of the same assessment instrument for comparing inter and intra-speakers is fundamental. This can be explained not only by the lack of evidence of agreement among the measurement methods available, but by the evidence that some variables can interfere in such measurements, such as the task and speech stimulus, presentation mode, type of response required and listener experience with the speaker.

As described in previous reports, it is likely that no single intelligibility measurement alone can be universally applied to all patients with speech disturbances, across several degrees of severity, and addressing all clinical and research objectives(8,10). Thus, independently of the method chosen by the clinician or researcher for assessing intelligibility, the interpretation of scores obtained should take into account all the confounding variables known to date, in a bid to prevent inappropriate generalizations of results for real communication situations.

#### References

1. Kent RD, Weismer G, Kent JF, Rosenbek JC. Toward phonetic intelligibility testing in dysarthria. *J Speech Hear Disord.* 1989;54:482-99.
2. Hustad KC. A closer look at transcription intelligibility for speakers with dysarthria: evaluation of scoring paradigms and linguistic errors made by listeners. *Am J Speech Lang Pathol.* 2006;15:268-77.
3. Hustad KC. Effects of speech stimuli and dysarthria severity on intelligibility scores and listener confidence ratings for speakers with cerebral palsy. *Folia Phoniatr Logop.* 2007;59: 306-17.
4. Sittler RW, Schiavetti N, Metz DE. Contextual effects in the measurement of hearing-impaired speakers' intelligibility. *J Speech Hear Res.* 1983;26:30-4.

5. Kempler D, Van Lancker D. Effect of speech task on intelligibility in dysarthria: a case study of Parkinson's Disease. *Brain Lang.* 2002;80:449-64.
6. Tikofsky RS, Tikofsky RP. Intelligibility measures of dysarthric speech. *J Speech Hear Res.* 1964;7:325-33.
7. Yorkston KM, Beukelman DR. A comparison of techniques for measuring intelligibility of dysarthric speech. *J Commun Disord.* 1978;11:499-512.
8. Beukelman DR, Yorkston KM. The relationship between information transfer and speech intelligibility of dysarthric speakers. *J Commun Disord.* 1979;12:189-96.
9. McGarr NS. The effect of context on the intelligibility of hearing and deaf children's speech. *Lang Speech.* 1981;24:255-64.
10. Yorkston KM, Beukelman DR. Communication efficiency of dysarthric speakers as measured by sentence intelligibility and speaking rate. *J Speech Hear Disord.* 1981;46:296-301.
11. Monsen RB. The oral speech intelligibility of hearing-impaired talkers. *J Speech Hear Disord.* 1983;48:286-96.
12. Ellis LW, Fucci DJ. Magnitude-estimation scaling of speech intelligibility: effects of listeners' experience and semantic-syntactic context. *Percept Mot Skills.* 1991;73:295-305.
13. Garcia JM, Cannito MP. Influence of verbal and nonverbal contexts on the sentence intelligibility of a speaker with dysarthria. *J Speech Hear Res.* 1996;39:750-60.
14. Bain C, Ferguson A, Mathisen B. Effectiveness of the speech enhancer on intelligibility: a case study. *J Med Speech-Lang Pathol.* 2005;13(2):85-95.
15. Hubbard DJ, Kushner D. A comparison of speech intelligibility between esophageal and normal speakers via three modes of presentation. *J Speech Hear Res.* 1980;23:909-16.
16. Garcia JM, Crowe LK, Redler D, Hustad K. Effects of spontaneous gestures on comprehension and intelligibility of dysarthric speech: a case report. *J Med Speech-Lang Pathol.* 2004;12(4):145-8.
17. Hustad KC. Estimating the intelligibility of speakers with dysarthria. *Folia Phoniatr Logop.* 2006;58(3):217-28.
18. Prins D, Bloomer H. Consonant intelligibility: a procedure for evaluating speech in oral cleft subjects. *J Speech Hear Res.* 1968;11:128-37.
19. Ellis L, Reynolds L, Fucci D, Benjamin B. Effects of gender on listeners' judgments of speech intelligibility. *Percept Mot Skills.* 1996;83:771-5.
20. Finizia C, Lindstrom J, Dotevall H. Intelligibility and perceptual ratings after treatment for laryngeal cancer: laryngectomy versus radiotherapy. *Laryngoscope.* 1998;108(1):138-43.
21. Liss JM, Stephanie MS, Caviness JN, Adler C. The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *J Acoust Soc Am.* 2002;112(6):3022-30.
22. Weismer G, Laures JS. Direct magnitude estimates of speech intelligibility in dysarthria: effects of a chosen standard. *J Speech Lang Hear Res.* 2002;45:421-34.
23. Porter KF, Bradley S. A comparison of three speech intelligibility measures for deaf students. *Am Ann Deaf.* 1985;130:514-25.
24. Furia CLB, Kowalski LP, Latorre MRDO, Angelis EC, Martins NMS, Barros APB, Karina CB. Speech intelligibility after glossectomy and speech rehabilitation. *Arch Otolaryngol Head Neck Surg.* 2001;127:877-83.
25. Zraick RI, Dennie TM, Tabbal SD, Hutton TJ, Hicks GM, O'Sullivan PS. Reliability of speech intelligibility ratings using the Unified Parkinson Disease Rating Scale. *J Med Speech-Lang Pathol.* 2003;11(4):227-40.
26. Zraick RI, Davenport DJ, Tabbal SD, Hutton TJ, Hicks GM, Patterson JH. Reliability of speech intelligibility ratings using the Unified Huntington Disease Rating Scale. *J Med Speech-Lang Pathol.* 2004;12(1):31-40.
27. Hanson EK, Beukelman DR. Effect of omitted cues on alphabet supplemented speech intelligibility. *J Med Speech-Lang Pathol.* 2006;14(3):185-96.
28. Hustad KC. Influence of alphabet cues on listeners ability to identify sound segments in sentences produced by speakers with moderate and severe dysarthria. *J Med Speech-Lang Pathol.* 2006;14(4):249-52.
29. Whitehill TL, Wong CC-Y. Contributing factors to listener effort for dysarthric speech. *J Med Speech-Lang Pathol.* 2006;14(4):335-41.
30. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307-10.