

# Inteligibilidade: efeitos da análise de transcrição e do estímulo de fala\*\*\*

## Intelligibility: effects of transcription analysis and speech stimulus

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

---

\*Fonoaudióloga. Doutoranda em Distúrbios da Comunicação Humana pela Universidade Federal de São Paulo (Unifesp). 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 Unifesp. Professor Adjunto do Departamento de Fonoaudiologia da Unifesp.

\*\*\*Trabalho Realizado no Núcleo de Investigação Fonoaudiológica em Neuropsicolinguística da Unifesp.

Artigo Original de Pesquisa

Artigo Submetido a Avaliação por Pares

Conflito de Interesse: não

Recebido em 01.05.2009.  
Revisado em 26.12.2009; 07.02.2010.  
Aceito para Publicação em 22.04.2010.

### Abstract

Background: intelligibility measures are limited to providing information on the severity level of clinical cases. A key limitation is that such measures are sensitive to changes in performance only in subjects with a determined severity level of speech disturbance. Aim: to investigate the influence of stimuli type and transcription analysis on intelligibility measures of speakers with no communication disorders. Method: an experimental study with no intervention procedures was developed. Two groups of subjects with no communication disorders took part in the research. The group of speakers was composed by 30 adults. Speech samples were recorded by repeating three lists of stimuli (sentences, words and non-words) equally distributed according to parameters of frequency of phonemes, syllabic structures and word length. The group of listeners was composed by 60 young adults who orthographically transcribed the speech samples. Two transcription intelligibility measures were obtained for each list of stimuli: percentage of correct answers per syllable unit and per item (for each sentence, word and non-word). Results: intelligibility scores were statistically higher for syllable units than for the other items. Regarding intelligibility scores per syllables, a statistical difference was observed amongst scores for sentences, words and non-words. Conclusion: both transcription analysis and stimulus type influenced the intelligibility scores of the studied population, especially when non-words were used as speech material. The handling of these variables can help to improve intelligibility tests.

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

### Resumo

Tema: apesar de seu amplo emprego com fins clínicos e de pesquisa, as medidas de inteligibilidade da fala por transcrição são criticadas por fornecer apenas informações sobre o grau de severidade dos quadros, bem como por ter sensibilidade restrita, dependendo do grau de alteração do paciente. Objetivo: investigar a influência da análise de transcrição e do tipo de estímulo sobre as medidas de inteligibilidade de sujeitos sem distúrbios da comunicação. Método: um estudo experimental sem intervenção foi realizado. Dois grupos de sujeitos sem distúrbios da comunicação participaram desta pesquisa. O grupo de falantes foi composto por 30 adultos. Amostras de fala foram gravadas em áudio a partir da repetição de três listas de estímulos (frases, palavras e pseudopalavras) igualmente distribuídas de acordo com os parâmetros: frequência dos fonemas, estruturas silábicas e extensão das palavras. O grupo ouvinte foi formado por 60 adultos jovens que transcreveram ortograficamente as amostras. Duas medidas de inteligibilidade foram obtidas para cada lista de estímulos: percentagem de respostas corretas por unidade silábica e por item (cada frase, palavra ou pseudopalavra). Resultados: os escores de inteligibilidade por unidade silábica foram estatisticamente superiores aos escores de inteligibilidade por item. Diferenças também foram observadas entre os escores de inteligibilidade por sílabas para frases, palavras e pseudopalavras. Conclusão: tanto a análise de transcrição quanto o tipo de estímulo influenciaram os escores da população estudada, especialmente quando as pseudopalavras foram utilizadas como material de fala. A manipulação destas variáveis pode ser útil ao aprimoramento dos testes de inteligibilidade da fala.

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

Referenciar este material como:



Barreto SS, Ortiz KZ. Intelligibility: effects of transcription analysis and speech stimulus (original title: Inteligibilidade: efeitos da análise de transcrição e do estímulo de fala). *Pró-Fono Revista de Atualização Científica*. 2010 abr-jun;22(2):125-30.

## Introduction

Several methods of assessing speech intelligibility have been developed, the majority of which are based on listeners transcribing speech samples and subsequent calculation of intelligibility scores<sup>1-12</sup>. However, these speech intelligibility scores are limited to providing information on degree of severity of clinical cases<sup>13,14</sup>. A further key limitation is that intelligibility measures are sensitive to changes in performance only in subjects with a given degree of severity of speech disturbance<sup>15</sup>. Given the crucial role of intelligibility in the rehabilitation of subjects with speech disorders, improving these measures is extremely important to clinical speech therapy practice. Exploiting the insights gained into the effects of certain variables, particularly those pertaining to the assessment instrument, may aid this process of refinement.

Although there is no compelling evidence that the type of transcription analysis, in terms of scoring paradigm, impacts intelligibility scores from a clinical viewpoint<sup>9</sup>, other as yet uninvestigated aspects related to this variable may influence speech intelligibility, such as transcription scoring level (syllable, word or sentence). Typically, when sentences or words are used as speech stimuli, correct answers are attributed to correctly transcribed words. However, in this case problems in only part of the word lead to rejection of the whole item, thus precluding more accurate identification of the unintelligible segments of the phonological sequence. Scoring of the transcription by syllabic unit may represent a viable alternative.

One such variable warranting special focus is the type of speech stimulus used. Results of studies involving individuals who are hearing-impaired and those with dysarthria have demonstrated that listeners make use of semantic cues in order to compensate for the deficits in the acoustic information of the altered speech<sup>16</sup>, according to the degree of severity of disorder<sup>1,11,15,17</sup>. Therefore, the semantic cues inherent to different types of speech stimuli may exert a particular influence on sensitivity of the intelligibility test. In particular, introduction of pseudowords as speech stimuli may constitute a more appropriate complementary measure for identifying deficits in speakers with mild dysarthria in view of the low sensitivity of intelligibility measures for usual sentences and words.

Against this background, the present study aimed to investigate the effects of transcription

analysis type (correct answers according to syllabic unit or to item) and the effects of stimuli type (sentences, words and pseudowords) on intelligibility scores of speakers with no communication disorders.

## Methods

An experimental study without intervention was carried out. This study was approved by the Research Ethics Committee of the Federal University of São Paulo (UNIFESP/ Parecer 0708/06).

### Participants

Two groups of subjects with no communication disorders participated in this study: a speaker group and a listener group. Speakers without communication disorders were studied because their intelligibility scores tend to be very similar to scores of individuals with mild dysarthria. Therefore, the findings in the population with no communication disturbances can be used to form a hypothesis on the performance of subjects with mild dysarthria.

The group of speakers comprised 30 adults of both genders whose Mother tongue was Brazilian Portuguese. Exclusion criteria adopted were: history of present or previous communication disorders, history of neurologic compromise (traumatic brain injury, stroke, epilepsy, among others), uncontrolled high blood pressure, and use of psychotropic medication or psychiatric history. Speakers presented a mean age of 40.4 years (SD= 13.2) and gender ratio of 1:1.

The listener group included 60 subjects who spoke Brazilian Portuguese as their Mother tongue (age: mean  $\pm$  SD = 22.4  $\pm$  4.2 years). Exclusion criteria applied were: history of language, learning or cognitive disorders, hearing loss on basic audiologic test, and familiarity with speakers or with the stimuli employed in the intelligibility assessment. These factors were controlled because they could have interfered in intelligibility measurement.

### Material

Three lists of stimuli were used, namely: sentences, words and pseudowords. The three stimulus types differed in terms of quantity of semantic information that can be inferred by the listener and exploited to assist in the task of

decoding speech samples. A list of phonetically balanced phrases comprising 25 short sentences was used. The sentences contained an average of five words and nine syllables, with 520 phonetic occurrences and 237 syllables (appendix 1)18.

The list of words and pseudowords was devised based on the frequency of phonemes, types of syllabic structure and word lengths found in the list of sentences. The pseudowords were constructed based on words with the same length and syllable structures as the words from the sentences in which one to three phonemes were altered. The degree of correlation of lists devised and the list of sentences calculated using Spearman's correlation coefficient ( $r$ ) proved extremely strong for all parameters considered ( $r = 0.993$  and  $p < 0.001$ ). Word and pseudoword lists were identical to each other in terms of spread of parameters, containing 60 stimuli, each with 260 phonetic occurrences and 118 syllables (appendix 2).

The following equipment was used to record the speech samples: a Cyber Acoustics, model AC-100 microphone headset, a Toshiba, model Satellite L25 Notebook and the Sound Forge 4.5 program (Sony Creative Software Inc, Madison, WI, USA). The Praat 4.4.13 program along with model CD-6631MV Edifier head phones were also used for sound file editing and the transcription task.

## Procedures

### Recording and editing of speech samples

The speakers were instructed to repeat the three lists of stimuli at a natural speed and intensity. Verbal repetition was preferred over reading to prevent any affect of speaker's reading ability on performance. The list order was counterbalanced in the group of speakers to prevent an ordering effect on results. The recording was carried out in a silent environment, with the subject seated and microphone placed 5 cm from their mouth. Original sound files were edited into 145 individual files per speaker for later presentation to listeners.

### Transcription task

The listeners performed orthographic transcription of the speech sample. Each listener was randomly designated to transcribe the sample of one speaker only in a bid to minimize the effect of prior knowledge of stimuli on test results. The speech sample of each speaker was transcribed by

two listeners in order to minimize the influence of variability of listeners on intelligibility scores. The order of presentation of the list of stimuli followed the original recording order, whereby items from each list were presented once only, one by one and at intervals dictated by the listener's transcription pace. All listeners set the volume to a comfortable level which was subsequently used throughout the transcription task.

### Transcription Analysis and Scoring

Two types of analysis were employed according to the transcription scoring level: correct syllabic units and correct items. For syllabic unit transcription analysis, each correctly decoded syllable was scored while for analysis by item, scores were assigned per sentence, word or pseudoword. Intelligibility was measured by the percentage of correctly transcribed syllables or items for the list of stimulus. Transcribed stimuli were considered correct when phonemic correspondence was observed between the orthographic transcription and the expected production of the target stimuli by speakers. Since the samples of speakers were transcribed by two listeners, the end scores of each assessed subject were calculated based on the average of the scores attributed by the respective listeners.

### Statistical Analysis

Differences among means of continuous data were assessed using parametric (Student's  $t$ ) and non-parametric (Wilcoxon) tests, showing similar results in all cases. Only parametric test results are shown. Assessment agreement among intelligibility scores was ascertained by calculating the limits of agreement proposed by Bland and Altman (1986). A probability ( $p$ ) of less than 0.05 was considered statistically significant. All tests were two-tailed. Ninety five percent confidence intervals (CI) were calculated for differences between means and for intra-class correlation coefficient (ICC). All analyses were performed using version 11.5.1 of the SPSS (Statistical Package for the Social Sciences) statistical package for Windows (SPSS Inc, 2002).

### Reliability

Reliability of interlistener and intralister intelligibility scores was analyzed. The intra-class correlation coefficient was used to check interlistener agreement (two listeners per speaker). Strong agreement was found for interlistener scores (sentences: ICC= 0,85; 95% CI= 0,70 to 0,92; words:

ICC= 0,71; 95%CI= 0,48 to 0,85; and pseudowords: ICC= 0,84; 95%CI= 0,70 to 0,92).

In order to ascertain intralistener reliability, approximately 10% of listeners, selected randomly, were given a second transcription task for the same speaker, which they had seen two weeks earlier under test conditions. Means obtained in the first and second assessments were compared using the Student t test for paired samples. No differences were observed in listener scores on test-retest (sentences:  $t(6) = -1,2$ , 95%CI= -1,6 to 0,5;  $p = 0,253$ ; words:  $t(6) = -1,6$ , 95%CI= -6,5 to 1,4;  $p = 0,166$ ; and pseudowords:  $t(6) = -1,9$ , 95%CI= -5,0 to 0,6;  $p = 0,104$ ).

## Results

Speech intelligibility scores by type of transcription analysis and stimulus are shown in Table 1.

The analysis of transcription by syllabic unit yielded statistically higher scores per item across all three stimulus types as shown in Table 2. Despite the differences detected, the limits of agreement indicate that the measurements show a clinically relevant difference only when pseudowords were used as stimuli.

Regarding measurement of intelligibility by syllables, intelligibility scores of sentences, words and pseudowords differed significantly, with higher scores for sentences and lower score for pseudowords (Table 2). Similarly, the limits of agreement revealed a clinically relevant difference between means in comparisons of pseudowords with other types of stimulus.

Plots of differences between intelligibility scores versus their respective means revealed for all comparisons made that the lower the mean intelligibility, the greater the discrepancy among scores. Graph 1 depicts one of these comparisons (Bland and Altman graph).

TABLE 1. Descriptive analysis of intelligibility scores.

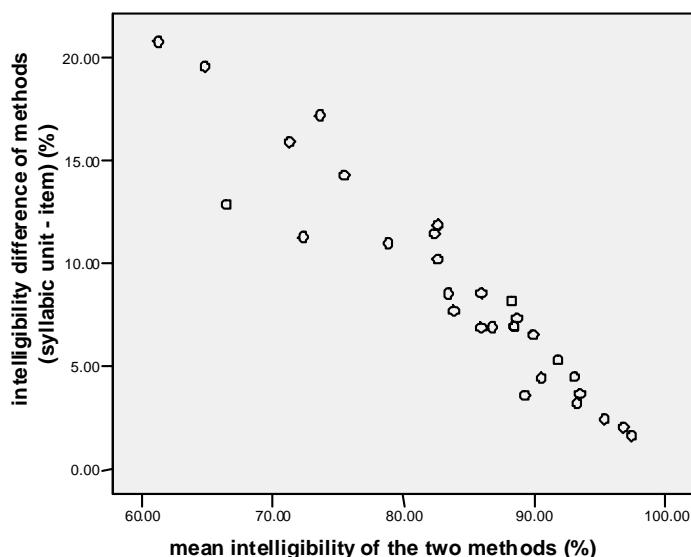
Intelligibility Scores	Mean (%)	Standard Deviation	Median (%)	Minimum (%)	Maximum (%)
item - sentences	97,60	3,38	100,00	88,00	100,00
item - words	95,00	4,25	96,25	85,83	100,00
item - pseudowords	79,72	12,10	82,92	50,83	96,66
syllable - sentences	99,28	1,09	99,79	95,57	100,00
syllable - words	96,90	2,99	97,67	88,13	100,00
syllable - pseudowords	88,46	7,29	90,25	71,61	98,30

TABLE 2. Comparison among intelligibility scores by type of transcription analysis and stimulus.

Comparison	t(29)	95%CI	p	Limits of Agreement
<b>type of transcription analysis</b>				
SEN= INT / item X INT / syllable	3,55	0,71 to 2,65	=0,001*	-3,42 to 6,78
WOR= INT / item X INT / syllable	6,70	1,32 to 2,49	<0,001*	-1,16 to 4,96
PSEUDO= INT / item X INT / syllable	6,70	6,84 to 10,64	<0,001*	-1,25 to 18,70
<b>type of stimulus (INT / syllable)</b>				
SEN X WOR	6,22	1,60 to 3,16	<0,001*	-1,72 to 6,48
SEN X PSEUDO	9,14	8,40 to 13,25	<0,001*	-1,90 to 23,54
WOR X PSEUDO	8,55	6,42 to 10,47	<0,001*	-2,15 to 19,05

SEN: sentences; WOR: words; PSEUDO: pseudowords; INT/item: intelligibility scores by item; INT/syllable: intelligibility scores by syllable; \* $p < 0,05$

GRAPH 1. Difference among intelligibility scores of pseudowords analyzed by syllabic unit and item, plotted against mean intelligibility scores.



## Discussion

With regard to the influence of type of transcription analysis, independently of stimulus type, scores obtained from analysis of transcription by syllable were greater than those by item. These higher scores may be explained by the difference in accuracy of the error analysis seen when the same sample was analyzed at different levels (syllable versus item). In view of the number of phonetic occurrences which make up the corpus of each list, analysis by item led to greater losses, since adequately identified occurrences were not scored, given rejection of the whole item upon identification of only partial errors.

Despite the significant difference observed among these intelligibility scores, only pseudoword differences led to disagreement among intelligibility measures. According to limits of agreement (-1,25 and 18,70), the variation in differences observable for this stimulus type indicates that, from a clinical viewpoint, these scores furnish distinct information regarding intelligibility of the subjects assessed.

No studies addressing the influence of level of transcription analysis on measures of intelligibility of speech were found in the literature.

Regarding type of stimulus, the intelligibility of sentences attained higher scores than all other forms, while intelligibility of words was higher than pseudowords, indicating that the more linguistic information made available to the listener, the greater the intelligibility of speech scores. The findings of previous studies confirm this evidence in subjects

with no communication disorders<sup>16</sup>, in speakers with dysarthria<sup>1,11,15</sup> and in those with hearing loss<sup>16,17</sup> for whom intelligibility of words in sentences attained higher scores than for intelligibility of words in isolation among speakers with milder cases. No other studies employing pseudowords were found.

In the present study, pseudowords were present in both stimulus type comparisons in which intelligibility of speech was affected and led to disagreement between the two measures (sentences versus pseudowords, and words versus pseudowords). Based on these results, we may infer that the absence of semantic cues interferes more in the assessment of intelligibility than mere reduction of these cues, at least amongst subjects whose speech attained high levels of intelligibility. Amongst intelligible speakers, the absence of linguistic cues increases the sensitivity of the test for minimal losses of speech signal (acoustic-phonetic information). In the case of sentences and words, such losses are easily compensated by semantic information inferred by listeners from the speech material.

The trend of increased difference among less intelligible subjects observed for all comparisons made (as shown in Graph 1), suggests that even greater differences are likely to be observed on assessments of dysarthric speakers. Thus, a greater influence of transcription analysis and stimulus type on intelligibility score is likely in this population.

## Conclusion

Drawing on the analysis of the results obtained in the present study, we may conclude that both transcription analysis and stimulus types influenced the intelligibility scores of the population studied. A greater discrepancy was found for pseudowords. The findings suggest that stronger influences are likely to be found in speakers with speech disturbances, where the manipulation of these variables may be useful to help refine intelligibility tests.

Considering intelligibility of speech as a measure of quantity of information transferred, use of pseudowords in conjunction with transcription analysis based on scoring by syllables may be considered incoherent, since this group is devoid of semantic structure. Nonetheless, these measures may serve to complement speech intelligibility assessment of individuals with dysarthria by aiding identification of the speech production problems compromising intelligibility. In addition, these measures can increase sensitivity for speakers with mild alterations thus furthering understanding of this human communication disorder.

### Appendix 1

#### List of sentences<sup>18</sup>

- |                                     |                                       |   |
|-------------------------------------|---------------------------------------|---|
| 1. Não posso perder o ônibus.       | 9. Esqueci de pagar a conta.          | 17. Guardei o livro na primeira gaveta. |
| 2. Vamos tomar um cafezinho.        | 10. Os preços subiram ontem.          | 18. Hoje é meu dia de sorte.            |
| 3. Preciso ir ao médico.            | 11. O jantar está na mesa.            | 19. O sol está muito quente.            |
| 4. A porta da frente está aberta.   | 12. As crianças estão brincando.      | 20. Sua mãe acabou de sair de carro.    |
| 5. A comida tinha muito sal.        | 13. Choveu muito nesse fim de semana. | 21. Ela vai viajar nas férias.          |
| 6. Cheguei atrasado para a reunião. | 14. Estou morrendo de saudade.        | 22. Não quero perder o avião.           |
| 7. Vamos conversar lá na sala.      | 15. Olhe bem ao atravessar a rua.     | 23. Eu não conheci sua filha.           |
| 8. Depois liga para mim.            | 16. Preciso pensar com calma.         | 24. Ela precisa esperar na fila.        |
|                                     |                                       | 25. O banco fechou sua conta.           |



Appendix

List of words

1. Ai	21. Isso	41. Irmão
2. Ar	22. Lua	42. Comeu
3. Se	23. Tia	43. Peguei
4. No	24. Amor	44. Gostou
5. Só	25. Assim	45. Cuidar
6. Já	26. Onze	46. Quartos
7. Vi	27. Este	47. Errada
8. Fé	28. Suor	48. Exame
9. Pá	29. Faca	49. Piada
10. Te	30. Linha	50. Acerta
11. Som	31. Nasce	51. Esquina
12. Vim	32. Pena	52. Altura
13. Cão	33. Morro	53. Acordou
14. Deu	34. Será	54. Noventa
15. Vai	35. Tombo	55. Certeza
16. Fez	36. Mundo	56. Trabalha
17. Mil	37. Mexer	57. Criando
18. Pães	38. Secas	58. Prendendo
19. Era	39. Pêras	59. Opinião
20. Até	40. Livre	60. Abraçado

List of pseudowords

1. Iu	21. Oma	41. Arção
2. Er	22. Dua	42. Mosseu
3. Che	23. Nia	43. Lequei
4. Ne	24. Ezar	44. Pestar
5. Sá	25. Anãe	45. Suidem
6. Rá	26. Esto	46. Quama
7. Gue	27. Saór	47. Asselha
8. Lé	28. Fopa	48. Afirda
9. Qui	29. Benha	49. Avérta
10. Su	30. Dezo	50. Miaços
11. Bim	31. Pema	51. Apatou
12. Vom	32. Terro	52. Escoça
13. Fão	33. Sira	53. Almore
14. Tou	34. Conge	54. Mureza
15. Lai	35. Sunto	55. Covendo
16. Tós	36. Cares	56. Bremida
17. Dal	37. Neços	57. Criante
18. Peis	38. Matir	58. Pringuendo
19. Éda	39. Ompar	59. Ocanião
20. Ipé	40. Tivra	60. Atravade

References

1. Yorkston KM, Beukelman DR. A comparison of techniques for measuring intelligibility of dysarthric speech. *J Commun Disord.* 1978;11:499-512.
2. 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.
3. Hustad KC, Jones T, Dailey S. Implementing speech supplementation strategies: effects on intelligibility and speech rate of individuals with chronic severe dysarthria. *J Speech Lang Hear Res.* 2003;46(2):462-74.
4. 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.
5. Hustad KC, Beukelman DR. Effects of linguistic cues and stimulus cohesion on intelligibility of severely dysarthric speech. *J Speech Lang Hear Res.* 2001;44:497-510.
6. 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.
7. Hanson EK, Beukelman DR. Effect of omitted cues on alphabet supplemented speech intelligibility. *J Med Speech-Lang Pathol.* 2006;14(3):185-96.
8. Hustad KC. Estimating the intelligibility of speakers with dysarthria. *Folia Phoniatri Logop.* 2006;58(3):217-28.
9. 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.
10. Whitehill TL, Wong CCY. Contributing factors to listener effort for dysarthric speech. *J Med Speech-Lang Pathol.* 2006;14(4):335-41.
11. Hustad KC. Effects of speech stimuli and dysarthria severity on intelligibility scores and listener confidence

ratings for speakers with cerebral palsy. *Folia Phoniatri Logop.* 2007;59: 306-17.

12. Hustad KC. The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *J Speech Lang Hear Res.* 2008;51:562-73.

13. Kent RD, Weismer G, Kent JF, Rosenbek JC. Toward phonetic intelligibility testing in dysarthria. *J Speech Hear Disord.* 1989;54:482-99.

14. Dykstra AD, Hakel ME, Adams SG. Application of the ICF in reduced speech intelligibility in dysarthria. *Semin Speech Lang.* 2007;28(4):301-11.

15. 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.

16. McGarr NS. The effect of context on the intelligibility of hearing and deaf children's speech. *Lang Speech.* 1981;24:255-64.

17. Sittler RW, Schiavetti N, Metz DE. Contextual effects in the measurement of hearing-impaired speakers' intelligibility. *J Speech Hear Res.* 1983;26:30-5.

18. Costa MJ, Iorio MCM, Mangabeira-Albernaz PL. Reconhecimento de fala: desenvolvimento de uma lista de sentenças em português. *Acta Awho.* 1997;16(4):164-73.

19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307-10.