Speech intelligibility measurement

A latent variable approach on utterances' transcriptions

Jose Rivera[1], Sven de Maeyer[2], and Steven Gillis[3]

[1] Department of Training and Education Sciences,
University of Antwerp, Antwerp, Belgium
E-mail: JoseManuel.RiveraEspejo@uantwerpen.be
(corresponding author)

[2] Department of Training and Education Sciences,
University of Antwerp, Antwerp, Belgium
E-mail: sven.demaeyer@uantwerpen.be

[3] Computational Linguistics, and Psycholinguistics Research Centre
University of Antwerp, Antwerp, Belgium
E-mail: steven.gillis@uantwerpen.be

August 9, 2022

Abstract

Contents

List of Figures

List of Tables

1 Introduction

Intelligible speech can be defined as the extent in which the elements in a speaker's acoustic signal, e.g. phonemes or words, can be correctly recovered by a listener [9, 11, 18, 19]. Intelligible spoken language carries an important societal value, as its attainment requires all core components of speech perception, cognitive processing, linguistic knowledge, and articulation to be mastered [9]. In that sense, *speech intelligibility* is considered a milestone in children's language development, and more practically, it is qualified as the ultimate checkpoint for the success of speech therapy, and the 'gold standard' for assessing the benefit of cochlear implantation [4].

Multiple approaches can be taken to quantify (measure) *speech intelligibility* [1, 2, 7, 10], but among them, *objective rating* methods on stimuli recovered from spontaneous speech tasks have received special attention [2, 10]. In objective rating methods, listeners transcribe children's utterances orthographically (or phonetically), and use such information to construct an intelligibility score. The construction of the score can be done in many ways, e.g. counting and normalize the number of (un)intelligible syllables or words in the utterances [7, 12], or calculating the transcriptions' entropy, a measure that expresses the degree of (dis)agreement in the data [2, 16]. In that sense, the method tries to infer intelligibility from the extent in which a set of transcribers, can identify the words contained in multiple utterances [2].

As the literature suggests, objective rating procedures produce more valid[1] and reliable[2] scores than any other available procedure [2, 6], as the method does not hinge in the use or production of a *subjective rating scale*, i.e. a scale based on a personal perception of the child's intelligibility. Moreover, the previous advantages are further emphasized by the use of stimuli gathered from spontaneous speech tasks, as they have a greater level of ecological validity, especially compared to contextualized utterances or reading at loud tasks [7, 5].

However, although the literature is clear on the method's benefits to measure *speech intelligibility* [1, 2, 10], we notice the statistical approaches used to model such data still face three important issues, and these come to the detriment of the measurement procedure's sophistication.

First, as previous paragraphs reveal, the intelligibility scores are 'complex' in nature, however, such 'complexity' is rarely fully considered in the statistical modeling procedure. The problem with the later is that, because the data does not fulfill the typical assumptions, e.g. normality, it might lead us to erroneous conclusions [citation]. On the one hand, outcomes such as the number of (un)intelligible words are discrete, while the entropy scores are continuous in nature. In addition, there is the consideration that both measures are constraint in specific bounds, i.e. the (normalized) number of (un)intelligible words cannot be negative, while the entropy scores are in the bounds between zero and one. Finally, given the *objective rating* procedure nature, the scores are produced in a clustered manner, i.e. we observe several score measurement per child. Considering all of the above, it is clear the modeling requires some adjustments to account for all of these nuances in the data.

The statistical procedures applied in the literature have always assumed 'normality' [2, 8, 10], and some papers even used multilevel modeling to deal with the clustered nature of the data [2], however, to the authors knowledge no paper have dealt with all of the data 'complexity' at once. This leads us to believe that there are gains by using more sophisticated statistical models, more specifically, by using generalized linear mixed models (GLMM) [3, 13, 15].

[1] validity is understood as the extent to which scores are appropriate for their intended interpretation and use [14, 17].

[2] reliability is though as the extend to which a measure would give us the same result over and over again [17], i.e. measure something, free from error, in a consistent way.

Bibliography

[1] Boonen, N., Kloots, H. and Gillis, S. [2020]. Rating the overall speech quality of hearing-impaired children by means of comparative judgements, *Journal of Communication Disorders* **83**: 1675–1687.
doi: https://doi.org/10.1016/j.jcomdis.2019.105969.

[2] Boonen, N., Kloots, H., Nurzia, P. and Gillis, S. [2021]. Spontaneous speech intelligibility: early cochlear implanted children versus their normally hearing peers at seven years of age, *Journal of Child Language* pp. 1–26.
doi: https://doi.org/10.1017/S0305000921000714.

[3] Breslow, N. and Clayton, D. [1993]. Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association* **88**(421): 9–25.
doi: https://doi.org/10.2307/2290687.
url: http://www.jstor.org/stable/2290687.

[4] Chin, S., Bergeson, T. and Phan, J. [2012]. Speech intelligibility and prosody production in children with cochlear implants, *Journal of Communication Disorders* **45**: 355–366.
doi: https://doi.org/10.1016/j.jcomdis.2012.05.003.

[5] Ertmer, D. [2011]. Assessing speech intelligibility in children with hearing loss: Toward revitalizing a valuable clinical tool, *Language, Speech, and Hearing Services in Schools* **42**(1): 52–58.
doi: https://doi.org/10.1044/0161-1461(2010/09-0081).

[6] Faes, J., De Maeyer, S. and Gillis, S. [2021]. Speech intelligibility of children with an auditory brainstem implant: a triple-case study, pp. 1–50. (submitted).

[7] Flipsen, P. [2006]. Measuring the intelligibility of conversational speech in children, *Clinical Linguistics & Phonetics* **20**(4): 303–312.
doi: https://doi.org/10.1080/02699200400024863.

[8] Flipsen, P. and Colvard, L. [2006]. Intelligibility of conversational speech produced by children with cochlear implants, *Journal of Communication Disorders* **39**(2): 93–108.
doi: https://doi.org/10.1016/j.jcomdis.2005.11.001.
url: https://www.sciencedirect.com/science/article/pii/S0021992405000614.

[9] Freeman, V., Pisoni, D., Kronenberger, W. and Castellanos, I. [2017]. Speech intelligibility and psychosocial functioning in deaf children and teens with cochlear implants, *Journal of Deaf Studies and Deaf Education* **22**(3): 278–289.
doi: https://doi.org/10.1093/deafed/enx001.

[10] Hustad, K., Mahr, T., Natzke, P. and Rathouz, P. [2020]. Development of speech intelligibility between 30 and 47 months in typically developing children: A cross-sectional study of growth, *Journal of Speech, Language, and Hearing Research* **63**(6): 1675–1687.
doi: https://doi.org/10.1044/2020$_{J}SLHR - 20 - 00008$.
url :$https: //pubs.asha.org/doi/abs/$10.1044/2020$_{J}SLHR - 20 - 00008$.

[11] Kent, R., Weismer, G., Kent, J. and Rosenbek, J. [1989]. Toward phonetic intelligibility testing in dysarthria, *Journal of Speech and Hearing Disorders* **54**(4): 482–499.
doi: https://doi.org/10.1044/jshd.5404.482.

[12] Lagerberg, T., Asberg, J., Hartelius, L. and Persson, C. [2014]. Assessment of intelligibility using childrens spontaneous speech: Methodological aspects, *International Journal of Language and Communication Disorders* **49**: 228–239.
doi: https://doi.org/10.1111/1460-6984.12067.

[13] Lee, Y. and Nelder, J. A. [1996]. Hierarchical generalized linear models, *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(4): 619–656.
doi: https://doi.org/10.1111/j.2517-6161.1996.tb02105.x.
url: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02105.x.

[14] Lesterhuis, M. [2018]. *The validity of comparative judgement for assessing text quality: An assessors perspective*, PhD thesis, University of Antwerp.

[15] McCullagh, P. and Nelder, J. [1983]. *Generalized Linear Models*, Monographs on Statistics and Applied Probability, Routledge.
doi: https://doi.org/10.1201/9780203753736.

[16] Shannon, C. [1948]. A mathematical theory of communication, *The Bell System Technical Journal* **27**(3): 379–423.
doi: https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[17] Trochim, W. [2022]. The research methods knowledge base.
url: https://conjointly.com/kb/.

[18] van Heuven, V. [2008]. Making sense of strange sounds: (mutual) intelligibility of related language varieties. a review, *International Journal of Humanities and Arts Computing* **2**(1-2): 39–62.
doi: https://doi.org/10.3366/E1753854809000305.

[19] Whitehill, T. and Chau, C. [2004]. Single-word intelligibility in speakers with repaired cleft palate, *Clinical Linguistics and Phonetics* **18**: 341–355.
doi: https://doi.org/10.1080/02699200410001663344.