

Speech intelligibility measurement

A latent variable approach on utterances' transcriptions

Jose Rivera¹, Sven de Maeyer², and Steven Gillis³

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 1, 2022

Abstract

Contents

1	Introduction	4
	Bibliography	6

List of Figures

List of Tables

1 Introduction

Intelligible speech can be defined as the extent to which the elements in an speaker’s acoustic signal, e.g. phonemes or words, can be correctly recovered by a listener [14, 22, 21, 11]. 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 [11]. 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 *speech intelligibility* [1, 2, 10, 13], and among them, *objective rating* methods [13] on spontaneous speech tasks have received a especial attention. In objective rating methods, listeners transcribe children’s utterances orthographically or phonetically, and use such information to construct an intelligibility score. The stimuli that listeners hear, and later transcribe, are recovered from children’s spontaneous speech tasks. In that sense, the method infers intelligibility, from the extent in which a set of transcribers can identify the word contained in an utterance [2].

As the literature suggest, objective rating methods are considered to produce more valid and reliable scores than any other method [2, 8], as the rating procedure does not hinges in a subjective rating scale, i.e. a scale based on a personal perception of the intelligibility, but rather in actual transcriptions of the stimuli. Here validity is understood as the extent to which scores are appropriate for their intended interpretation and use [15, 20], while reliability is though as the extend to which a measure would give us the same result over and over again [20], i.e. measure something, free from error, in a consistent way. Moreover, it is though that stimuli coming from spontaneous speech tasks has the highest level of ecological validity [10, 6], followed by contextualized utterances and reading at loud tasks.

Although the literature conceptualize *speech intelligibility* as in the previous paragraph [2], and further suggest multiple perspectives and approaches to measure it [1, 2, 13], the statistical approach used to test reasearch hypothesis to authors opinion is lacking of ... knowledge no paper tries to estimate the individuals’ intelligibility score. Moreover, no paper test their research hypothesis directly on an intelligibility outcome, but rather on surrogate measures of it. For instance, Flipsen [10] uses

This is exacerbated

We believe this paper make three specific contributions to the understanding of the factors that drive the intelligibility of spoken language. First, we develop a novel analysis using a latent variable approach [7]. More specifically, we model *speech intelligibility* as a latent variable that can be inferred from the entropy replicates. This method offers three specific benefits. On the one hand, the method ‘constructs’ an intelligibility score, which in turn, allow us to test different hypothesis and even make individual comparisons at the appropriate level. On the other hand, it allow us to control for different sources of variation. This is particularly important as, by failing to account for the appropriate hierarchies in the data, we could be ‘manufacturing’ false confidence in the parameter estimates, leading us to incorrect inferences [16]. Finally, the method also provides a ‘criterion’ on how reliable are the entropy replicates to measure speech intelligibility.

Second, we use Directed Acyclic Graph (DAG) [18, 5] to depict all the relevant variables though to influence speech intelligibility. We describe in detail our causal and non-causal hypothesis, and supplement our description with a causal diagram. The benefit of the method lies, not only, in that it makes the assumptions of our hypothesis more transparent, but also allow us to derive statistical procedures from the aforementioned causal assumptions [16, 24, 19].

Accompanying the intelligibility assessment methods, the literature supply a myriad of factors that are thought also contribute to the (under)development of intelligible spoken language [17, 3, 12, 9]. Among these are audiology related factors, such chronological age, age at implantation, the duration of device use, *hearing age*, bilateral or contralateral cochlear implantation, and the children’s preoperative and postoperative hearing levels. On the other hand, there are also child related factors, such as the cause of the hearing impairment (genetic, infections), additional disabilities (mental retardation, speech motor problems), and gender. Finally, there are also environmental factors, such as communication modality.

Third and final, we wrap the analysis procedure under the Bayesian framework, providing the assumptions, and the steps required to reproduce the computational implementation of the models.

Considering all of the above, this paper seeks to investigates the speech intelligibility levels of normal hearing (NH) versus hearing-impaired children with cochlear implants (HI/CI). For that purpose, ten utterances recordings, from thirty two NH and HI/CI children, were selected from a large corpus of *spontaneously spoken speech* collected by the CLiPS research center. Additionally, we set up an experiment,

where one hundred language students transcribed each stimuli to the Qualtrics environment [23]. Finally, the transcriptions were transformed into an entropy measure per utterance, which served as our outcome variable.

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] Boons, T., Brokx, J., Dhooge, I., Frijns, J., Peeraer, L., Vermeulen, A., Wouters, J. and van Wieringen, A. [2012]. Predictors of spoken language development following pediatric cochlear implantation, *Ear and Hearing* **33**(5): 617–639.
doi: <https://doi.org/10.1097/AUD.0b013e3182503e47>.
- [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] Cinelli, C., Forney, A. and Pearl, J. [2022]. A crash course in good and bad controls, *SSRN*.
doi: <http://dx.doi.org/10.2139/ssrn.3689437>.
url: <https://ssrn.com/abstract=3689437>.
- [6] 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\(2010/09-0081\)](https://doi.org/10.1044/0161-1461(2010/09-0081)).
- [7] Everitt, B. [1984]. *An Introduction to Latent Variable Models*, Monographs on Statistics and Applied Probability, Springer Dordrecht.
doi: <https://doi.org/10.1007/978-94-009-5564-6>.
- [8] 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).
- [9] Fagan, M., Eisenberg, L. and Johnson, K. [2020]. Investigating early pre-implant predictors of language and cognitive development in children with cochlear implants, in M. Marschark and H. Knoors (eds), *Oxford handbook of deaf studies in learning and cognition*, Oxford University Press, pp. 46–95.
doi: <https://doi.org/10.1093/oxfordhb/9780190054045.013.3>.
- [10] 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>.
- [11] 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>.
- [12] Gillis, S. [2018]. Speech and language in congenitally deaf children with a cochlear implant, in E. Dattner and D. Ravid (eds), *Handbook of Communication Disorders: Theoretical, Empirical, and Applied Linguistic Perspectives*, De Gruyter Mouton, chapter 37, pp. 765–792.
doi: <https://doi.org/10.1515/9781614514909-038>.
- [13] 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_JSLHR-20-00008.
url: https://pubs.asha.org/doi/abs/10.1044/2020_JSLHR-20-00008.

- [14] 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>.
- [15] Lesterhuis, M. [2018]. *The validity of comparative judgement for assessing text quality: An assessors perspective*, PhD thesis, University of Antwerp.
- [16] McElreath, R. [2020]. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*, Chapman and Hall/CRC.
- [17] Niparko, J., Tobey, E., Thal, D., Eisenberg, L., Wang, N., Quittner, A. and Fink, N. [2010]. Spoken Language Development in Children Following Cochlear Implantation, *JAMA* **303**(15): 1498–1506.
doi: <https://doi.org/10.1001/jama.2010.451>.
- [18] Pearl, J. [2009]. *Causality: Models, Reasoning and Inference*, Cambridge University Press.
- [19] Rohrer, J., Schmukle, S. and McElreath, R. [2021]. The only thing that can stop bad causal inference is good causal inference, *PsyArXiv* .
doi: <https://doi.org/10.31234/osf.io/mz5jx>.
- [20] Trochim, W. [2022]. The research methods knowledge base.
url: <https://conjointly.com/kb/>.
- [21] 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>.
- [22] 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>.
- [23] Wright, B. [2005]. Qualtrics. (Version December 2018).
url: www.qualtrics.com.
- [24] Yarkoni, T. [2020]. The generalizability crisis, *The Behavioral and brain sciences* **45**(e1).
doi: <https://doi.org/10.1017/S0140525X20001685>.