Speech intelligibility:

A generalized latent variable approach on utterances' entropies

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

November 13, 2022

Abstract

Intelligibility is defined as "the extent to which a speaker's message is understood by the listener" [14]. Its attainment carries an important societal value, as it is a milestone in language development [4]. For its measurement, orthographic transcriptions of utterances are used to construct an entropy score, which express the degree of (dis)agreement in such transcriptions [2, 21]. However, although the benefits of using transcriptions to (indirectly) quantify intelligibility are clear [1, 2, 7], the statistical procedures used to model such data are not as sophisticated as the measurement procedures.

Consequently, we propose a novel data analysis using a Bayesian implementation of the Generalized Linear Latent and Mixed Model (GLLAMM) [16, 18, 17, 19, 22]. The statistical procedure offers four benefits. First, it allows to model the bounded entropy data. Second, it 'constructs' a speaker's latent intelligibility scale. Third, it test our research hypothesis at the appropriate level. And fourth, it avoids 'manufacturing' false confidence in the parameter estimates, producing correct statistical inferences [11].

As a result we find that, not modeling the bounded nature of the data could lead us to an overestimation of the parameter estimates' precision. For our hypothesis, we see that HI/CI children with genetic etiology have similar levels of intelligibility as NH kids, at 'hearing ages' of five. However, children with other etiologies have a significantly lower levels of intelligibility, at same ages. Moreover, we find that NH children develop their intelligibility with each 'hearing year' at a higher rate than HI/CI kids, contrary to what was previously found [2].

Contents

List of Figures

List of Tables

1 Introduction

Intelligible spoken language requires all core components of speech perception, cognitive processing, linguistic knowledge, and articulation to be mastered [5]. In that sense, its attainment carries an important societal value, as it 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].

But what is speech *intelligibility*?. *Intelligibility* can be broadly defined as "the extent to which a speaker's message is actually understood by the listener" [14]. But in a more narrow sense, it refers to the listener's ability to successfully identify (decode) the words in a message [5, 9, 25, 27]. The latter definition is more helpful, as it sets a clear contrast with comprehensibility, which involves the listener's ability to understand the message, and its intent [14, 23].

However, indifferent of its broad or narrow definition, the literature reveal that intelligibility can be further compromised by features of the communicative environment, such as noise [12]; by features of the speaker, like speaking rate [13] or accent [8, 15]; or features of the listener, like vocabulary mastery [26]. Moreover, the latter emphasizes its highly dynamical nature, where changes in intelligibility stem from on line adaptations of the speaker, to the listener and/or the context.

Therefore, we can say that speech intelligibility generate considerable interest for its societal value, but its measurement pose interesting challenges, particularly because of its entanglement with other features of the communication.

Considering the previous, the literature suggest two perspectives from which intelligibility can be assessed: the message and listener's perspective [1, 2]. The first, also known as acoustic studies, is focused on assessing separately particular characteristics of the speech samples, e.g. their pitch, duration, stress, or the articulation of vowels and consonants [20]. Whereas the second, also known as perceptual studies, is centered on making holistic assessments of the speech stimuli, e.g. measure their overall quality [1, 2]. The former is justified by the fact that using the speech samples, we can detect articulatory, acoustic, and auditory characteristics of intelligible utterances. In contrast, the latter is justified by the notion that intelligibility is a concept 'that everyone can judge', but can only be measured indirectly [6, 24].

Focusing our attention on perceptual studies, 'objective' rating methods on children's utterances recovered from spontaneous speech tasks, have received special attention [2, 7]. In these methods, listeners transcribe children's utterances orthographically (or phonetically), and use these transcriptions as information to construct an intelligibility score; more precisely, an entropy score that expresses the degree of (dis)agreement in the transcriptions [2, 21]. As a result, we obtain scores that are clustered and bounded in nature. Clustered because we get multiple measurements per child (one per utterance), and bounded because their values are in the continuum between zero and one.

Therefore, 'objective' rating methods try to infer intelligibility from the extent in which a set of transcribers can identify the words contained in the utterances [2]. In other words, we get a 'proxy' measure of the speaker's intelligibility as judged by a listener, a snapshot of his/her performance under a specific set of circumstances [7]. Moreover, the epistemological certainty in such 'snapshot' as a measure of intelligibility stems from the design and steps taken to collect the data.

However, although the literature is clear on the benefits of 'objective' rating methods to (indirectly) quantify intelligibility [1, 2, 7], we notice the statistical procedures used to model such data are not at par of the measurement procedure's sophistication.

Previous research have considered the clustered nature of the data but ignored its bounded nature, where averaging was considered a 'valid' option for modeling [2]. We argue that the latter practice is not appropriate, as with bounded data not only the location (average), but also the spread (variance) of the distribution, might inform about the speaker's intelligibility [10].

Furthermore, in order to understand or intervene on the factors that drives speech intelligibility, first one needs to 'construct an error free' *intelligibility* scale [3], a characteristic not possessed by the entropy measures nor its averages.

Considering all of the above, we propose a novel analysis of the entropy data using a Bayesian implementation of the Generalized Linear Latent and Mixed Model (GLLAMM) [16, 18, 17, 19, 22]. The statistical procedure offers four benefits. First, it allows to appropriately model the bounded entropy data. Second, it provides a way to 'construct' the speaker's latent intelligibility scale. Third, it allow us to test our research hypothesis at the appropriate level. And fourth, as a result from the first two, we successfully avoid producing false confidence in the parameter estimates, which help us to produce correct statistical inferences [11].

We find the proposed method bring new insights about the use of replicated entropy scores to measure intelligibility, and on how some factors affect the (under)development of children's intelligibility.

On the one hand, the method reveal that, not integrating the bounded nature of the data in the modeling procedure could lead us to wrongful statistical conclusions. More precisely, it could lead us to an overestimation of the parameter estimates' precision.

Lastly, our hypothesis tests reveal that hearing impaired children with cochlear implants (HI/CI) and genetic etiology have similar levels of intelligibility as normal hearing kids (NH), when both groups have a 'hearing ages' of five. However, the same cannot be said for children with other etiologies, like CMV infection or other causes, as they start a significantly lower level of intelligibility at same 'ages'. Moreover, our tests found enough evidence to assert that NH children develop their intelligibility with each 'hearing year' at a higher rate than HI/CI kids. This offer evidence contrary to what was previously found [2].

Finally, we observe our results support the hypothesis that HI/CI children with severe hearing loss, as accounted by the pure tone average, develop their language at a slower rate than their NH counterparts.

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] Carroll, J. [2006]. *Measurement error in nonlinear models: a modern perspective*, Chapman and Hall/CRC.
doi: https://doi.org/10.1201/9781420010138.

[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] 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.

[6] Guilford, J. [1954]. *Psychometric methods*, McGraw-Hill Book Company.

[7] 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$.

[8] Jenkins, S. [2000]. Cultural and linguistic miscues: a case study of international teaching assistant and academic faculty miscommunication, *International Journal of Intercultural Relations* **24**(4): 477–501.
doi: https://doi.org/10.1016/S0147-1767(00)00011-0.
url: https://www.sciencedirect.com/science/article/pii/S0147176700000110.

[9] 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.

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

[11] McElreath, R. [2020]. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*, Chapman and Hall/CRC.

[12] Munro, M. [1998]. The effects of noise on the intelligibility of foreign-accented speech, *Studies in Second Language Acquisition* **20**(2): 139–154.
doi: https://doi.org/10.1017/S0272263198002022.

[13] Munro, M. and Derwing, T. [1998]. The effects of speaking rate on listener evaluations of native and foreign-accented speech, *Language Learning* **48**(2): 159–182.
doi: https://doi.org/10.1111/1467-9922.00038.
url: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9922.00038.

[14] Munro, M. and Tracey, D. [1999]. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners, *Language Learning* **49**(s1): 285–310.
doi: https://doi.org/10.1111/0023-8333.49.s1.8.
url: https://onlinelibrary.wiley.com/doi/abs/10.1111/0023-8333.49.s1.8.

[15] Ockey, G., Papageorgiou, S. and French, R. [2016]. Effects of strength of accent on an l2 interactive lecture listening comprehension test, *International Journal of Listening* **30**(1-2): 84–98.
doi: 1https://doi.org/0.1080/10904018.2015.1056877.

[16] Rabe-Hesketh, S., Skrondal, A. and Pickles, A. [2004a]. Generalized multilevel structural equation modeling, *Psychometrika* **69**(2): 167–190.
doi: https://www.doi.org/10.1007/BF02295939.

[17] Rabe-Hesketh, S., Skrondal, A. and Pickles, A. [2004b]. *GLLAMM Manual*, UC Berkeley Division of Biostatistics.
url: http://www.biostat.jhsph.edu/ fdominic/teaching/bio656/software-gllamm.manual.pdf.

[18] Rabe-Hesketh, S., Skrondal, A. and Pickles, A. [2004c]. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects, *Journal of Econometrics* **128**(2): 301–323.
doi: https://www.doi.org/10.1016/j.jeconom.2004.08.017.
url: http://www.sciencedirect.com/science/article/pii/S0304407604001599.

[19] Rabe-Hesketh, S., Skrondal, A. and Zheng, X. [2012]. Multilevel structural equation modeling, *in* R. Hoyle (ed.), *Handbook of Structural Equation Modeling*, The Guilford Press, chapter 30, pp. 512–531.

[20] Rowe, B. and Levine, D. [2018]. *A Concise Introduction to Linguistics*, Routledge.

[21] 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.

[22] Skrondal, A. and Rabe-Hesketh, S. [2004]. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*, Interdisciplinary Statistics, Chapman Hall/CRC Press.

[23] Smith, L. and Nelson, C. [1985]. International intelligibility of english: directions and resources, *World Englishes* **4**(3): 333–342.
doi: https://doi.org/10.1111/j.1467-971X.1985.tb00423.x.
url: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-971X.1985.tb00423.x.

[24] Stevens, S. [1946]. On the theory of scales of measurement, *Science* **103**(2684): 677–680.
doi: 10.1126/science.103.2684.677.
url: https://www.science.org/doi/abs/10.1126/science.103.2684.677.

[25] 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.

[26] Varonis, E. and Susan, G. [1985]. Non-native/non-native conversations: A model for negotiation of meaning, *Applied Linguistics* **6**(1): 71–90.
doi: https://doi.org/10.1093/applin/6.1.71.
url: https://academic.oup.com/applij/article-pdf/6/1/71/9741729/71.pdf.

[27] 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.