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
[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 21, 2022

Corresponding author: Jose Rivera, Department of Training and Education Sciences, University of Antwerp, Antwerp, Belgium. e-mail: JoseManuel.RiveraEspejo@uantwerpen.be

Finantial support: The project was financed by the Flemish Government through the Research Fund of the University of Antwerp (BOF).

Competing interests: The authors declare they have no conflict of interest regarding this manuscript.

Keywords— intelligibilty, children with coclear implants, utterances' entropy, generalized linear latent and mixed model.

Abstract

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 (Freeman et al.; 2017). 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 (Chin et al.; 2012).

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" (Munro and Tracey; 1999). But in a more narrow sense, it refers to the listener's ability to successfully identify or decode the words in a message (Freeman et al.; 2017; Kent et al.; 1989; van Heuven; 2008; Whitehill and Chau; 2004). 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 (Munro and Tracey; 1999; Smith and Nelson; 1985).

However, indifferent of its broad or narrow definition, the literature also reveal that intelligibility can be compromised by features of the communicative environment, such as noise (Munro; 1998); by features of the speaker, like speaking rate (Munro and Derwing; 1998) or accent (Jenkins; 2000; Ockey et al.; 2016); or features of the listener, like vocabulary mastery (Varonis and Susan; 1985). Moreover, all of the above further emphasizes its dynamic nature, where changes in intelligibility stem from online adaptations of the speaker to the listener and/or the context.

Considering the above, we can say that speech intelligibility generates considerable interest among researchers for its societal value, but its measurement pose interesting challenges.

In that sense, the literature suggests two perspectives from which intelligibility can be assessed: the message and listener's perspective (Boonen et al.; 2020, 2021). 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 (Rowe and Levine; 2018). Whereas the second, also known as perceptual studies, is centered on making holistic assessments of the speech stimuli, e.g. measuring their overall quality (Boonen et al.; 2020, 2021). The former is justified by the fact that by using speech samples we can detect articulatory, acoustic, and auditory characteristics of intelligible utterances. In contrast, the latter is justified on the fact that intelligibilty is an intuitively understood notion, i.e. "something" that anyone can judge, but because of its entanglement with other features of the communication, it can be measured only indirectly (Guilford; 1954; Stevens; 1946).

Focusing our attention on perceptual studies, "objective rating" methods on children's utterances, recovered from spontaneous speech tasks, have received special attention because of their high level of ecological validity (Boonen et al.; 2021; Ertmer; 2011; Flipsen; 2006; Hustad et al.; 2020). In these methods, listeners transcribe children's utterances orthographically (or phonetically), which later are used as information to construct an entropy score that expresses the degree of (dis)agreement in the transcriptions (Boonen et al.; 2021; Shannon; 1948). As a result, the scores are characterized by their clustered and bounded nature. The first is because the data have multiple measurements per child (one per utterance). While the second is because the score values are expresed in the continuum between zero and one.

Therefore, "objective rating" methods try to infer intelligibility from the extent to which a set of transcribers can identify the words contained in the utterances (Boonen et al.; 2021). In other words, the method gets 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 (Hustad et al.; 2020). 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 the aforementioned method to (indirectly) quantify intelligibility (Boonen et al.; 2020, 2021; Hustad et al.; 2020), 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 data clustering, but ignored its bounded nature, where averaging was considered a valid option for modeling (Boonen et al.; 2021). 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 entropies' distribution might inform about the speaker's intelligibility (McCullagh and Nelder; 1983). Therefore, considering a more sophisticated statistical procedure could improve our statistical inferences related to the intelligibility of individuals (McElreath; 2020).

The preceding statement is easier to understand with a thought experiment. Imagine three children with different patterns of entropy measures, all reporting the same mean entropy of 0.5. The patterns are: (a) scores closely agglomerated around 0.5, (b) scores loosely aglomerated around 0.5, and finally, (c) half of the scores agglomerated around 0.1 and the other half around 0.9. It is clear that from the mean score we can say that the three children have an "average" level of intelligibility. However, it is also clear from the spread of the scores that more uncertainty (to the asesssment of "average") should be assigned to child (c), followed by (b) and finally (a). This just mean that we can be more confident that child (a) has an "average" level of intelligibility, than in the other two cases, where (c) represent the extreme example of uncertainty. In that sense, we can easily notice that not only the average but also the spread of the entropies' distribution informs the level of intelligibility, and that we need a model that can integrate all these pieces of information coming from the data.

Furthermore, in order to understand or intervene on the factors that drives speech intelligibility, first one needs to 'construct an error free' *intelligibility* scale (Carroll; 2006), 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) (Rabe-Hesketh et al.; 2004a,c,b, 2012; Skrondal and Rabe-Hesketh; 2004). 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 (McElreath; 2020).

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.

Bibliography

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.

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.

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

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.

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).

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.

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.

Guilford, J. (1954). *Psychometric methods*, McGraw-Hill Book Company.

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.

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.

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.

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

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

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.

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.

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.

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: https://doi.org/0.1080/10904018.2015.1056877.

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.

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.

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.

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.

Rowe, B. and Levine, D. (2018). *A Concise Introduction to Linguistics*, Routledge.

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.

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

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.

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.

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.

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.

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.