Automatic assessment of transcript accuracy for speech intelligibility studies

Hans Rutger Bosker[1]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

HansRutger.Bosker@mpi.nl

In the field of speech perception, many studies assess the intelligibility of spoken stimuli by means of verbal repetition ('repeat back what you hear') or transcription tasks ('type out what you hear'). The intelligibility of a given stimulus is then often expressed in terms of percentage of words correctly reported from the target stimulus. Yet scoring the participants' raw transcripts for words correctly identified from the target stimulus is a time-consuming task, and hence resource-intensive. Moreover, there is no consensus on what protocol to use for the human scoring, limiting the reliability of human scores. The present paper evaluates various forms of 'fuzzy string matching' between participants' responses and target sentences as automated metrics of listener transcript accuracy. Fuzzy string matching is identified as a consistent, efficient, and accurate method for automated assessment of listener transcripts, as evidenced by high correlations with human-generated scores (highest r = 0.94) and a strong relationship to acoustic markers of speech intelligibility. Thus, fuzzy string matching provides a practical tool for speech scientists, allowing fast and reliable assessment of listener transcript accuracy in large-scale speech intelligibility studies.