Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 82 (2016) 26–37
www.elsevier.com/locate/specom

# Evaluating a distortion-weighted glimpsing metric for predicting binaural speech intelligibility in rooms

Yan Tang*, Richard J. Hughes, Bruno M. Fazenda, Trevor J. Cox

*Acoustics Research Centre, University of Salford, Salford M5 4WT, UK*

Received 20 December 2015; received in revised form 5 April 2016; accepted 14 April 2016
Available online 28 May 2016

**Abstract**

A distortion-weighted glimpse proportion metric (BiDWGP) for predicting binaural speech intelligibility were evaluated in simulated anechoic and reverberant conditions, with and without a noise masker. The predictive performance of BiDWGP was compared to four reference binaural intelligibility metrics, which were extended from the Speech Intelligibility Index (SII) and the Speech Transmission Index (STI). In the anechoic sound field, BiDWGP demonstrated high accuracy in predicting binaural intelligibility for individual maskers ($\rho \geq 0.95$) and across maskers ($\rho \geq 0.94$). The reference metrics however performed less well in across-masker prediction ($0.54 \leq \rho \leq 0.86$) despite reasonable accuracy for individual maskers. In reverberant rooms, BiDWGP was more stable in all test conditions ($\rho \geq 0.87$) than the reference metrics, which showed different predictive patterns: the binaural STIs were more robust for the stationary than for the fluctuating noise masker, whilst the binaural SII displayed the opposite behaviour. The study shows that the new BiDWGP metric can provide similar or even more robust predictive power than the current standard metrics.



*Keywords:* Speech; Binaural; Objective intelligibility metric; Glimpsing; Noise; Reverberation.

## 1. Introduction

Objective intelligibility measures (OIMs) provide a fast and robust approach to estimating the intelligibility of speech. They have therefore been widely adopted in place of subjective tests for interim intelligibility evaluation in fields in which speech intelligibility is a concern – such as in telephony quality assessment (ANSI S3.5, 1997; Fletcher, 1921), audiology for hearing impairment (Holube and Kollmeier, 1996; Santos et al., 2013), acoustics design (Houtgast and Steeneken, 1985; IEC, 2011) and algorithm development for speech enhancement and modification (Gomez et al., 2012; Taal et al., 2010). As the majority of the OIMs estimate intelligibility based on purely *monaural* listening, their usability may be limited in more practical situations in which listeners hear *binaurally*. Therefore, an added advantage of developing binaural OIMs is that the effects of room acoustics (e.g. reverberation) on how listeners hear sounds in realistic environments may be more accurately taken into account.

Nearly all existing binaural intelligibility metrics (e.g. Andersen et al., 2015; Beutelmann et al., 2010; Jelfs et al., 2011; Schlesinger et al., 2010; van Wijngaarden and Drullman, 2008; Zurek, 1993) extend their monaural counterpart such as the Speech Intelligibility Index (SII, ANSI S3.5, 1997), the Speech Transmission Index (STI, IEC, 2011) and the short-time objective intelligibility measure (Taal et al., 2010), by taking the head shadow effect and binaural interaction into account. As this study demonstrates, existing binaural metrics do not work reasonably well in all test conditions. More recently, Tang et al. (2015) proposed a method for predicting binaural speech intelligibility by extending the distortion-weighted glimpse proportion (DWGP, Tang, 2014). Originally developed as a monaural method, the DWGP metric provides an objective assessment of speech modification algorithms that aim to boost speech intelligibility in noise. In the binaural version of the DWGP metric (BiDWGP), the better ear effect resulted from the head-shadow effect is

* Corresponding author.
*E-mail address:* y.tang@salford.ac.uk (Y. Tang).

http://dx.doi.org/10.1016/j.specom.2016.04.003

modelled with better-ear glimpses, which are essentially the time-frequency regions of speech with energy exceeding the noise by a certain threshold. The binaural interaction is quantified using the binaural masking level difference (BMLD, Levitt and Rabiner, 1967). In Tang et al. (2015), evaluation using subjective listening tests in a simulated anechoic sound field demonstrated that the intelligibility predicted by BiDWGP is highly correlated with listener performance in a word identification task in both a single stationary or fluctuating noise masker (Pearson correlation coefficients $\rho_p = 0.98$), and two or three of the same types of masker ($\rho_p \geq 0.94$).

The monaural DWGP metric incorporates a distortion weighting factor with the glimpse proportion metric (GP, Cooke, 2006; Tang, 2014). This weighting factor was initially introduced in Tang (2014) to increase the consistency of predictions by the GP metric across different noise maskers, especially between stationary (e.g. speech-shaped noise) and fluctuating (e.g. single-talker competing speech) maskers (Tang et al., 2016). The calculation of the distortion weighting factor was inspired by a STI-based metric, the normalise-covariance metric (Holube and Kollmeier, 1996), which uses the cross-correlation coefficient of the reference clean and noise-corrupted speech envelopes within each frequency band to determine the speech-to-distortion level. The DWGP metric adopts this approach and uses the cross-correlation coefficient directly to weight the number of the glimpses in a frequency band. This enables DWGP to take into account the impact of the masker on the speech envelope, in addition to the masked-audibility that is accounted for by the original idea of glimpse detection. STI metrics are reported to perform well when predicting speech intelligibility in reverberation (e.g. Houtgast and Steeneken, 1985; Houtgast et al., 1980; Plomp et al., 1980). Thus, it may be hypothesised that the BiDWGP metric may also preserve its predictive power for reverberation conditions, as it contains a STI-inspired component which operates in the modulation domain. However, despite its accurate predictions in anechoic conditions (Tang et al., 2015), the performance of BiDWGP in reverberant conditions has never been assessed. Therefore, the main aim of this study is to explore whether and how well the BiDWGP metric can predict intelligibility in reverberation.

In Section 2 of this paper the BiDWGP metric and four other reference intelligibility metrics with their binaural extensions are introduced. To evaluate their performance, the model predictions are compared with subjective data obtained from two listening experiments conducted in a simulated anechoic sound field and three rooms varying in size and reverberation time, both with and without a noise masker (Section 3). In addition to speech-shaped noise, which has the long term average spectrum of the chosen speech corpus and is widely used in evaluations of objective intelligibility metrics, competing speech uttered by a female speaker was also tested as a masker. Predicting intelligibility in the presence of competing speech is challenging due to the large temporal fluctuations present in the competing speech, and the possibility of it introducing informational masking – thus, compared to speech-shaped noise, it is used less often as a masker in relevant studies. As listening to speech in the presence of other talkers is a common realistic scenario, examining the performance of predictors for competing speech maskers has practical implications. Section 4 focuses on discussing the aspects that affect the performance of the BiDWGP metric; its limitations and further work are also explored. Finally, we draw conclusions from the study in Section 5.

## 2. The distortion-weighted glimpse proportion metric and reference metrics

In this section a technical overview of the BiDWGP metric will be presented first, followed by introductions to four state-of-the-art metrics with their binaural extensions, including the binaural Speech Intelligibility Index, and the three binaural Speech Transmission Index metrics with different implementations. As each metric may take different inputs for analysis, for the sake of clarity, six variables are defined first, which will be further referred to in this section:

- $s$, $s'$: clean speech in anechoic and reverberant conditions
- $n$, $n'$: noise masker in anechoic and reverberant conditions
- $m$, $m'$: noise-corrupted speech (i.e. speech+noise mixture) in anechoic and reverberant conditions

### *2.1. An overview of the distortion-weighted glimpse proportion metric (BiDWGP)*

For anechoic conditions, Zurek (1993) suggested a method to estimate the effective binaural signals from a single channel signal using a free-field to eardrum transformation of the sound pressure level (Shaw and Vaillancourt, 1985). With this approach as the first stage of the BiDWGP metric, Tang et al. (2015) demonstrated that BiDWGP can predict binaural intelligibility well from just a set of single channel signals ($s$, $n$ and $m$), provided that the azimuth angle and distances for speech and masker sources relative to the listener are known. Further analyses have confirmed that intelligibility predictions by BiDWGP using single channel signals with the location information, and direct binaural signals are highly consistent ($\rho_p = 0.998$, and the Euclidean distance of 0.091 for indices falling between 0 and 1). However, the estimated binaural signals do not carry room acoustic information, and therefore cannot be used by the metric to account for the effects of room acoustics such as reverberation. In the current study, we assume binaural signals are available such that the estimation stage in Tang et al. (2015) is unnecessary.

The BiDWGP consists of two main components that account for the factors that negatively affect intelligibility: (1) masked-audibility due to energetic masking, and (2) distortion of the speech envelope due to temporal fluctuation and smearing.