Voice source features for forensic voice comparison – an evaluation of the GLOTTEX® software package: Errata and addenda

Ewald Enzinger[1,2], *Cuiling Zhang*[1,3], *Geoffrey Stewart Morrison*[1]

[1]Forensic Voice Comparison Laboratory, School of Electrical Engineering & Telecommunications, University of New South Wales, Australia
[2]Acoustics Research Institute, Austrian Academy of Sciences, Austria
[3]Department of Forensic Science & Technology, China Criminal Police University, Shenyang, China

e.enzinger@student.unsw.edu.au, cuiling-zhang@forensic-voice-comparison.net, geoff-morrison@forensic-voice-comparison.net

This documents includes errata and addenda to: Enzinger, E., Zhang, C., & Morrison, G.S. (2012). Voice source features for forensic voice comparison – an evaluation of the GLOTTEX® software package. In *Proceedings of Odyssey 2012: The Speaker and Language Recognition Workshop*. 25–28 June, Singapore, pp. 78–85.

The addenda include results from additional channel conditions not available at the time of submission of the published paper, but which did form part of the presentation given at the workshop. The errata are with respect to the results for the baseline system as well as systems fused with the baseline.

1. Errata

When using voice source features, recordings from session 1 of each speaker were used as nominal offender recordings. Inadvertently, when using MFCCs for the baseline system we used recordings from session 2 instead of recordings from session 1 as the offender recordings. The following results now consistently use session 1 of each speaker as the nominal offender recording. For the baseline system and for all systems fused with the baseline system the numeric values are different, however, there are no substantial differences in the pattern of the results. Corrected versions of figures 7, 9, 11 and 13 are provided below.

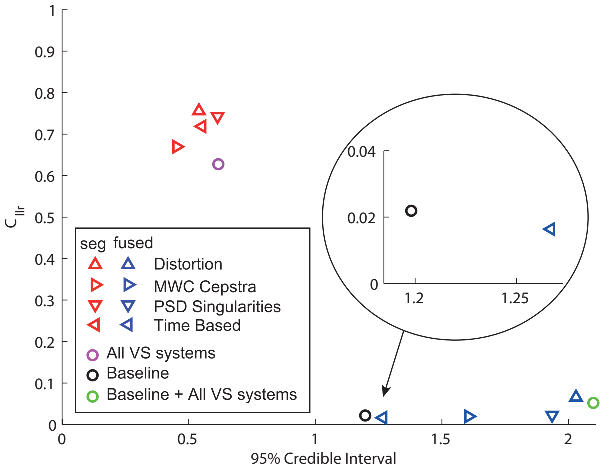

Figure 7: Measures for validity (C_{llr}) and reliability ($log10$ 95% credible interval) for the voice source feature systems applied to /n/ tokens individually (*red*) as well as after fusion with the generic fully-automatic baseline system (*blue*) (high-quality v high-quality recordings). Of the fusions of individual voice-source-feature systems with the baseline system, no fused system clearly outperformed the baseline system.

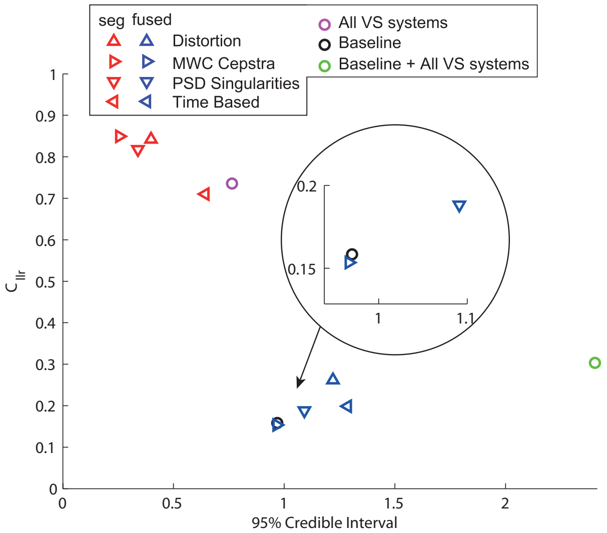

Figure 9: Measures for validity (C_{llr}) and reliability ($log10$ 95% credible interval) for the voice source feature systems applied to /n/ tokens individually (*red*) as well as after fusion with the generic fully-automatic baseline system (*blue*) (mobile-to-landline v mobile-to-landline recordings). The baseline system had a C_{llr} of 0.159 and a $\log_{10}$ 95% CI of 0.97. Of the fusions of individual voice-source-feature systems with the baseline system, no fused system clearly outperformed the baseline system.

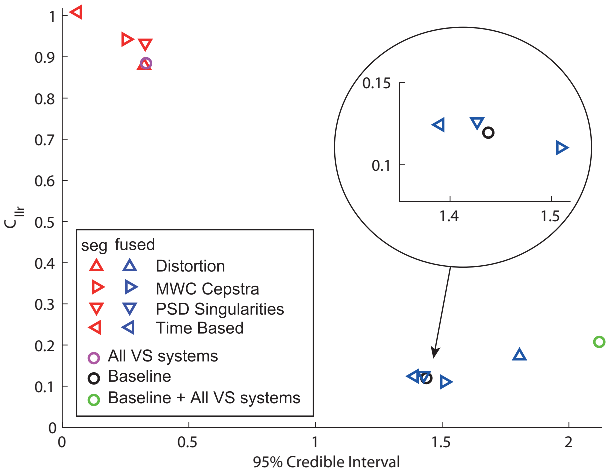

Figure 11: Measures for validity (C_{llr}) and reliability ($log10$ 95% credible interval) for the voice source feature systems applied to /n/ tokens individually (*red*) as well as after fusion with the generic fully-automatic baseline system (*blue*) (mobile-to-landline v high-quality recordings). The baseline system had a C_{llr} of 0.119 and a $\log_{10}$ 95% CI of 1.438. The best fused system was the *time-based* system with a C_{llr} of 0.124 and a $\log_{10}$ 95% CI of 1.39. It showed improvements in reliability at a loss in validity.

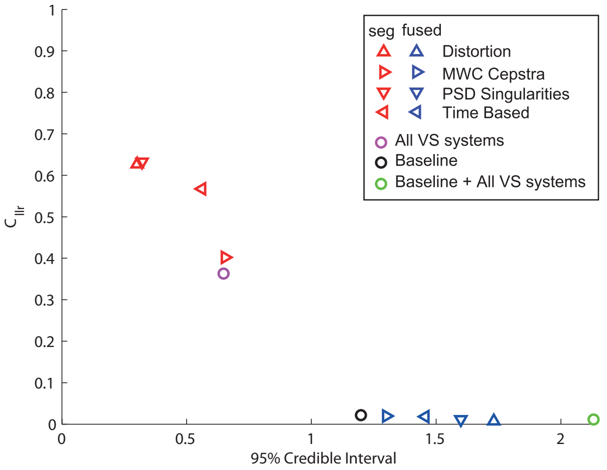

Figure 13: Measures for validity (C_{llr}) and reliability ($log10$ 95% credible interval) for the voice source feature systems applied to the total speech-active portion of the recordings individually (*red*) as well as after fusion with the generic fully-automatic baseline system (*blue*) (high-quality v high-quality recordings). Fusion of the individual systems with the baseline gave small improvements in validity with a loss in precision.

2. Addenda

Here are the plots for the new conditions. Adding the voice source feature based systems to the baseline system did not result in any substantial improvement in either condition.

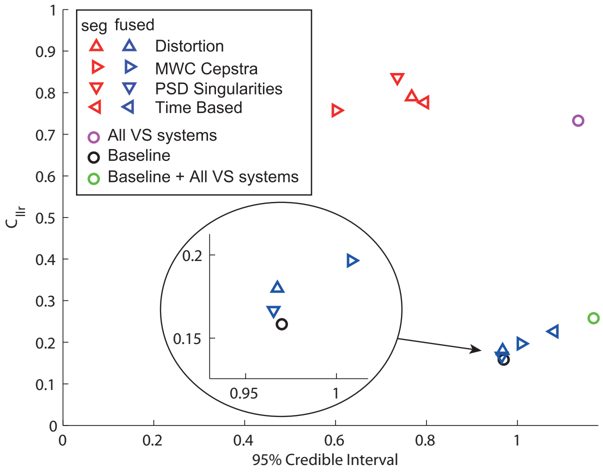

Figure 14: Measures for validity (C_{llr}) and reliability ($log10$ 95% credible interval) for the voice source feature systems applied to the total speech-active portion of the recordings individually (*red*) as well as after fusion with the generic fully-automatic baseline system (*blue*) (mobile-to-landline v mobile-to-landline recordings). Of the fusions of individual voice-source-feature systems with the baseline system, no fused system outperformed the baseline system.

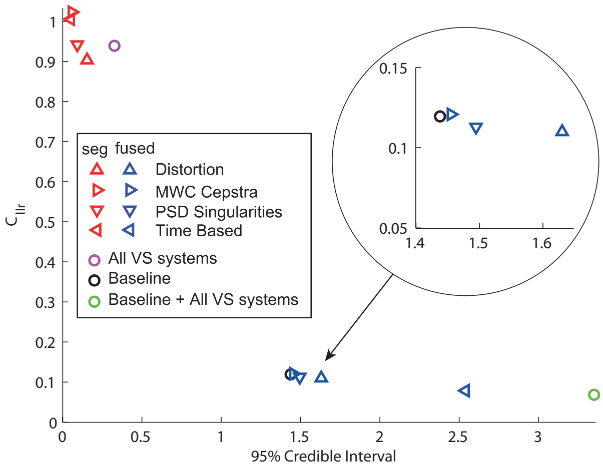

Figure 15: Measures for validity (C_{llr}) and reliability ($log10$ 95% credible interval) for the voice source feature systems applied to the total speech-active portion of the recordings individually (*red*) as well as after fusion with the generic fully-automatic baseline system (*blue*) (mobile-to-landline v high-quality recordings). Fusion of the individual systems with the baseline gave minor improvements in validity with a large loss in precision.