The VoicePrivacy 2020 Challenge

Odyssey 2020

Post evaluation analysis
Voice Similarity Matrices

Presenter: Paul-Gauthier Noé

Natalia Tomashenko [1]
Brij M.L. Srivastava [2]
Xin Wang [3]
Emmanuel Vincent [4]
Andreas Nautsch [5]
Junichi Yamagishi [3,6]
Nicholas Evans [5]
Jose Patino [5]
Jean-François Bonastre [1]
Paul-Gauthier Noé [1]
Massimiliano Todisco [5]
Mohamed Maouche [2]
Benjamin O'Brien [7]
Anais Chanclu [1]

1 LIA – University of Avignon – France
2 Inria – France
3 NII – Tokyo – Japan
4 Inria – France
5 Audio Security and Privacy Group, EURECOM – France
6 University of Edinburgh – UK
7 Aix-Marseille University – France

4th November 2020

Post-evaluation analysis: Voice Similarity Matrices

Speech Pseudonymisation Assessment Using Voice Similarity Matrices [Noe 2020]

Voice Similarity Matrices for the evaluation of:

- ✓ Differences in performance across speakers,
- ✓ Global De-Identification,
- ✓ Global Voice Distinctiveness Preservation.

Post-evaluation analysis: Voice Similarity Matrices

Voice Similarity Matrix:

$$M = \left(S(i,j)\right)_{1 \le i \le N, 1 \le j \le N}$$

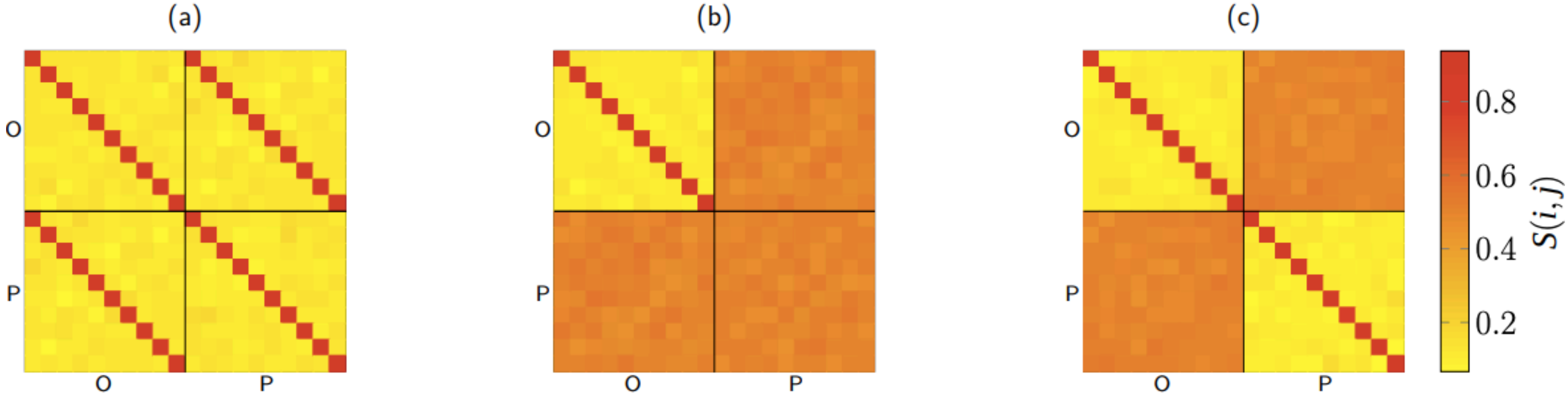
$$S(i,j) = sigmoid\left(\sum_{\substack{1 \le k \le n_i \\ 1 \le l \le n_j}} \frac{llr(x_k^{(i)}, x_l^{(j)})}{n_i n_j}\right)$$

where $x_q^{(p)}$ is the q-th segment of the p-th speaker, n_p is the number of segments from the p-th speaker and $llr(\cdot,\cdot)$ is the log likelihood-ratio score from the comparison of the two speech segments.

We build three voice similarity matrices:

- ✓ M_{OO} *within* the original set,
- ✓ M_{OP} *between* the original and pseudonymized sets,
- ✓ M_{PP} *within* the pseudonymised set.

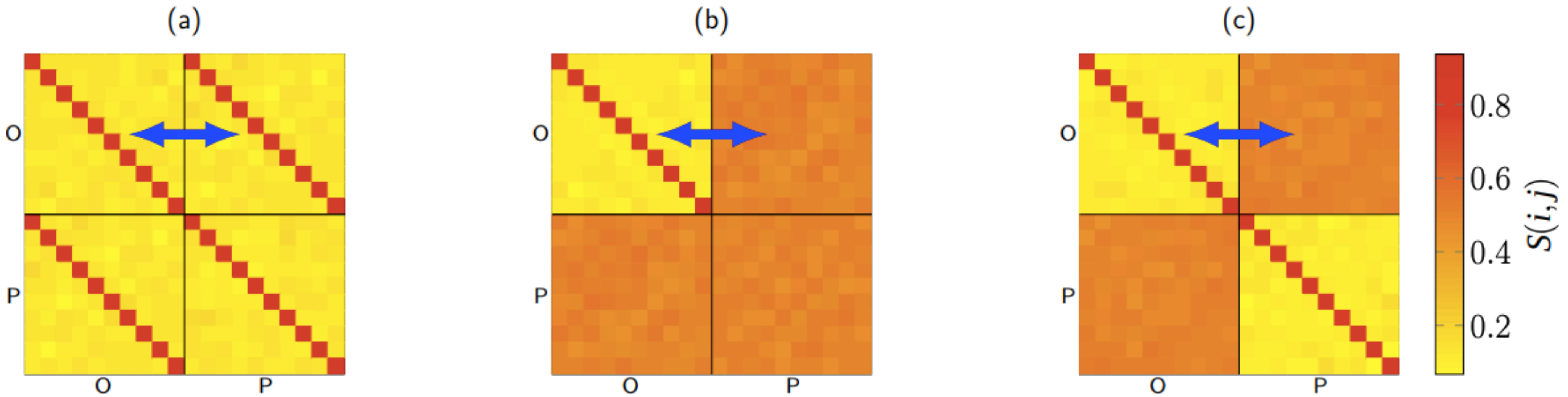
Post-evaluation analysis: Voice Similarity Matrices



Three artificial similarity matrices. The upper-left is M_{OO}, the upper-right and lower-left are M_{OP} and the lower-right is M_{PP}.

Diagonals comparison → Insight on the global performance

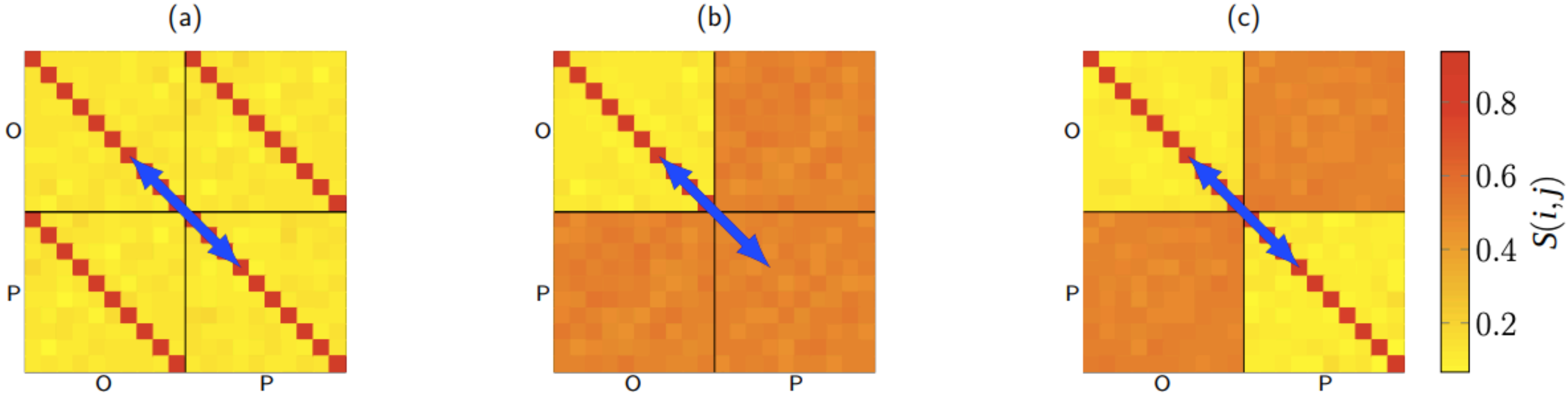
Post-evaluation analysis: Voice Similarity Matrices



Three artificial similarity matrices. The upper-left is M_{OO}, the upper-right and lower-left are M_{OP} and the lower-right is M_{PP}.

Diagonals comparison → Insight on the global performance

Post-evaluation analysis: Voice Similarity Matrices



Three artificial similarity matrices. The upper-left is M_{OO}, the upper-right and lower-left are M_{OP} and the lower-right is M_{PP}.

Diagonals comparison → Insight on the global performance

Post-evaluation analysis: Voice Similarity Matrices

We propose two metrics based on diagonals comparison:

De-Identification:

$$\mathrm{DeID} = 1 - \frac{D_{diag}(M_{OP})}{D_{diag}(M_{OO})}$$

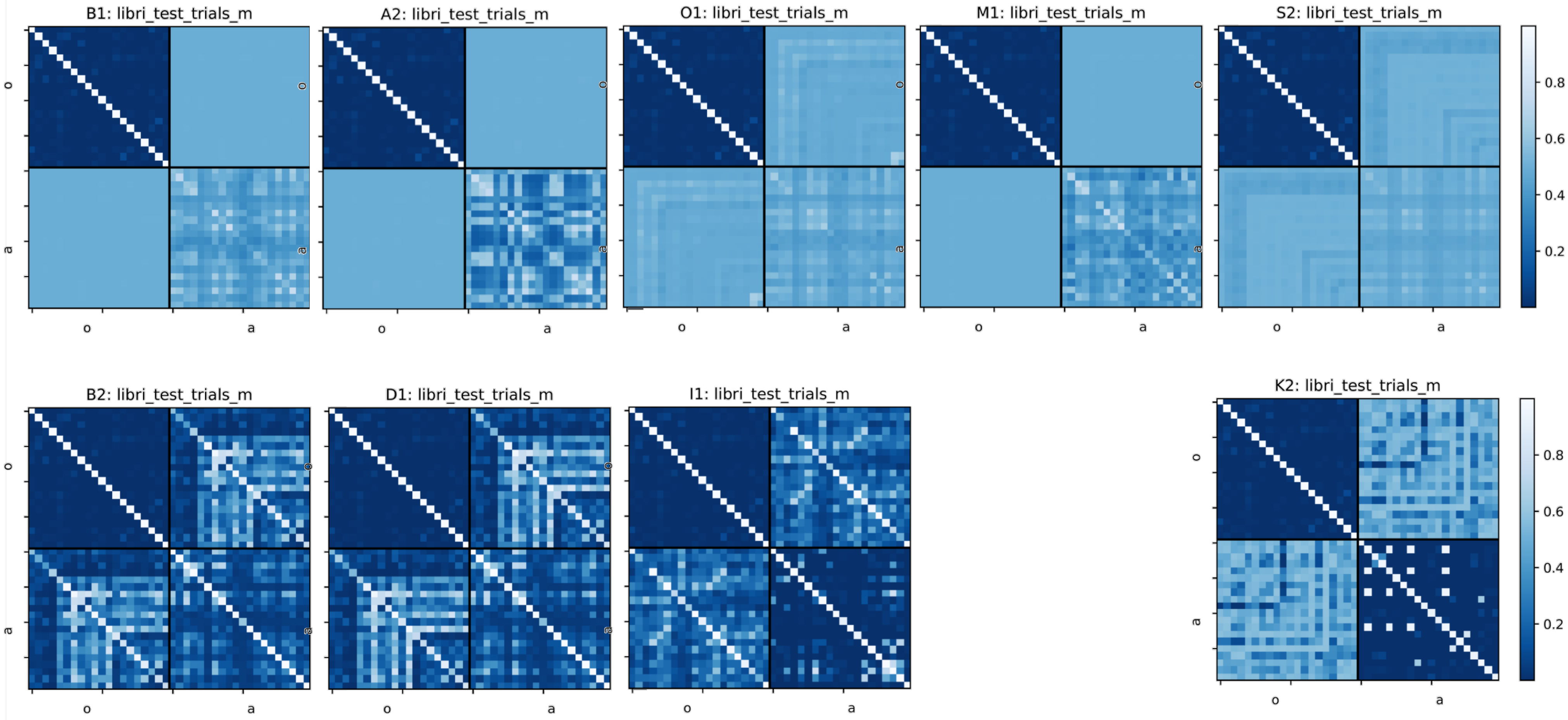
Voice Distinctiveness Preservation:

$$G_{VD} = 10\log_{10}\left(\frac{D_{diag}(M_{PP})}{D_{diag}(M_{OO})}\right)$$

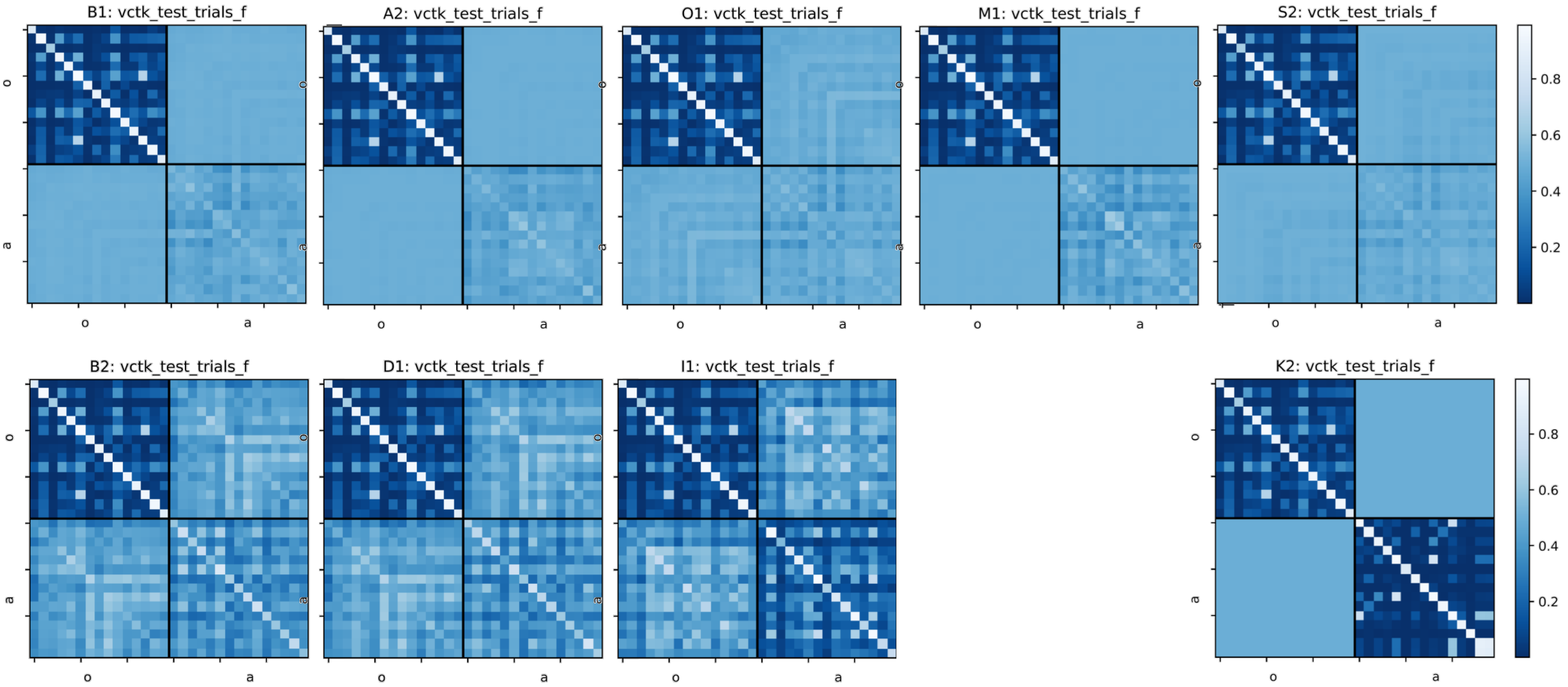
Where the diagonal dominance of a matrix M is defined as:

$$D_{diag}(M) = \left|\left(\sum_{1 \le i \le N} \frac{S(i,i)}{N}\right) - \left(\sum_{\substack{1 \le j \le N \\ 1 \le k \le N \\ j \ne k}} \frac{S(j,k)}{N(N-1)}\right)\right|$$

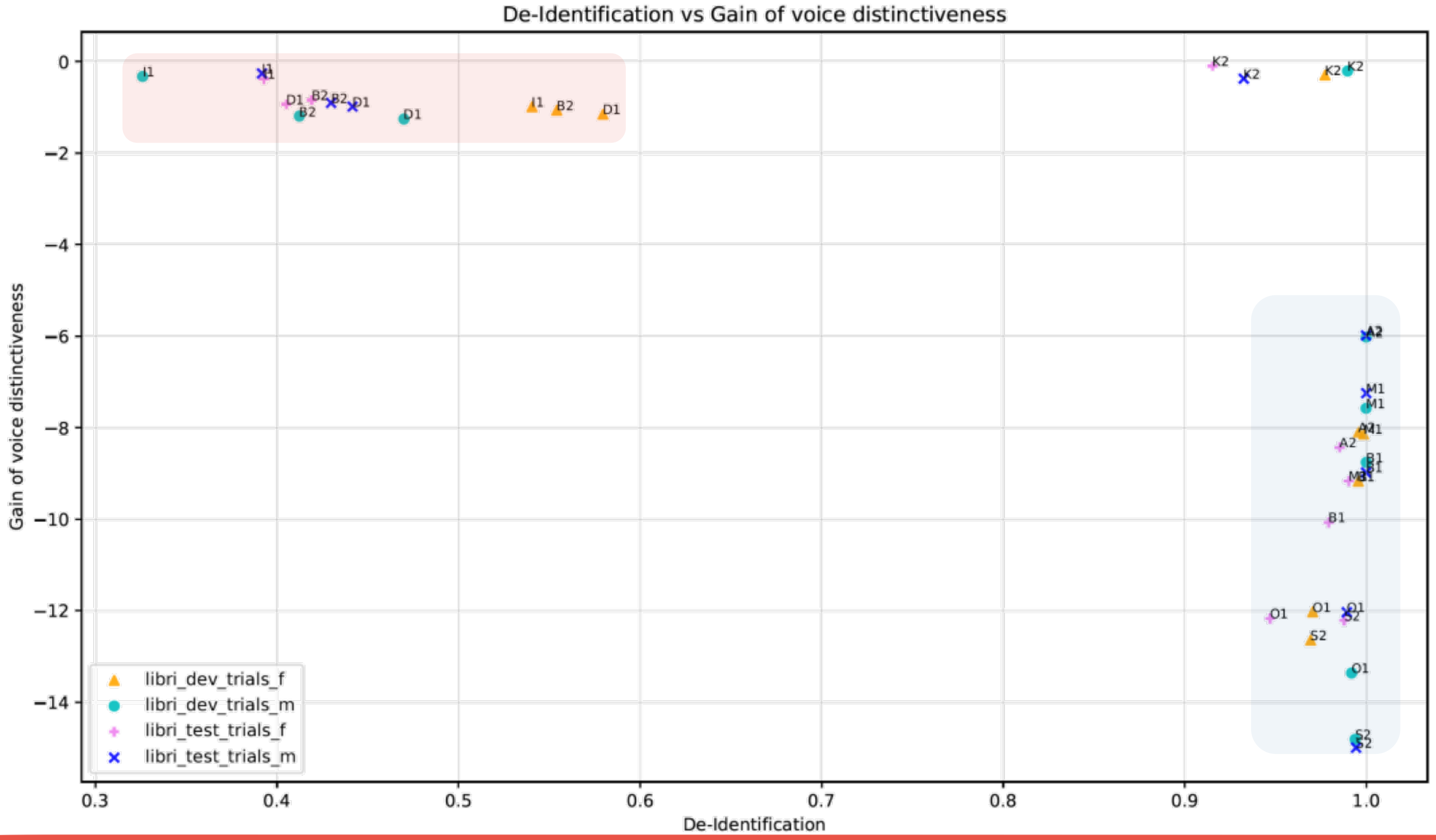
Similarity matrices: LibriSpeech-test-male



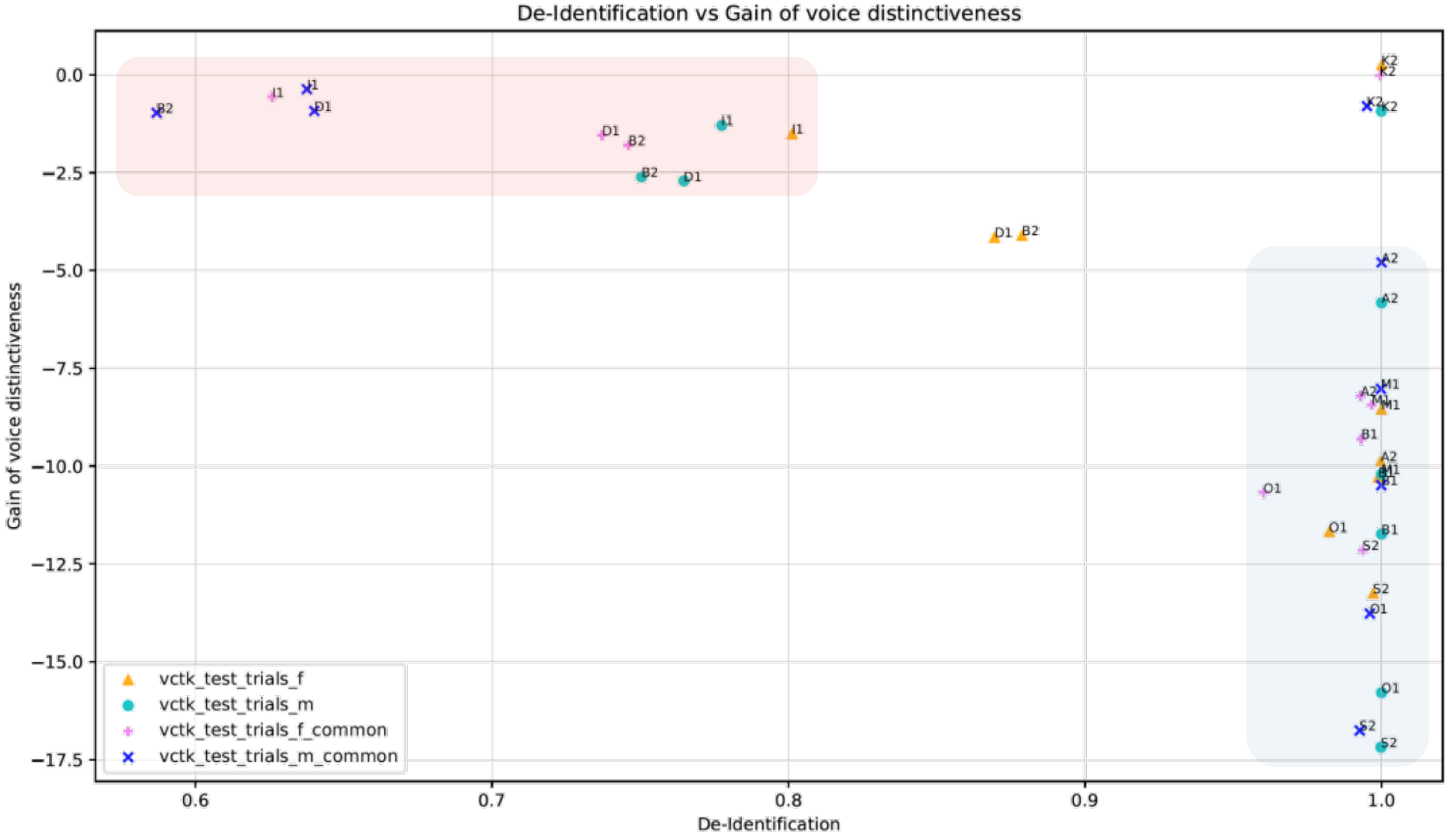
Similarity matrices: VCTK-test-female (different)



De-Identification & Gain of voice distinctiveness: LibriSpeech



De-Identification & Gain of voice distinctiveness: VCTK



Conclusions

Voice Similarity matrices

- ✓ Assess visually the De-Identification and Voice Distinctiveness,
- ✓ Differences of performance across speakers.

Two Metrics from the diagonals comparison

- ✓ Global De-Identification (DeID),
- ✓ Global Voice Distinctiveness Preservation (VDP),
- ✓ Most systems perform well either on DeID or on VDP but hardly on both.