**Supplementary material of the paper**

**J. Galić, Đ. Grozdić  
“Exploring the Impact of Data Augmentation Techniques on Automatic Speech Recognition System Development: A Comparative Study ”**

fig2.emf

Figure 2. Average recognition accuracy (in %) for three speech databases in HMM framework. The percent of the Whi-Spe subset used in training is given in abscissa.

HMM (Whi-Spe)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Speaker  Whi\_Spe subset | 2 | 6 | 7 | 8 | 9 | 14 | 15 | 16 | 18 | 19 |
| 10% | 80.33 | 98.67 | 94.67 | 100 | 98 | 98.33 | 97.67 | 99.67 | 96.33 | 92.33 |
| 20% | 88.67 | 99.67 | 97 | 100 | 99.67 | 99.33 | 97.33 | 100 | 97.67 | 92 |
| 30% | 90.33 | 99.33 | 98 | 100 | 99.67 | 99.67 | 97.67 | 100 | 98.67 | 94.67 |
| 40% | 89 | 100 | 97.67 | 100 | 99.33 | 99.67 | 97.67 | 100 | 98 | 94.67 |
| 50% | 91 | 99.67 | 97.67 | 100 | 99.67 | 99.67 | 97.33 | 100 | 98 | 95 |
| 60% | 91.33 | 99.67 | 97.33 | 100 | 99.67 | 99.67 | 97.33 | 100 | 98.67 | 95 |
| 70% | 90.33 | 100 | 98 | 100 | 100 | 99.67 | 97.33 | 100 | 98 | 96.33 |
| 80% | 90.33 | 100 | 98 | 100 | 100 | 99.67 | 97.33 | 100 | 98 | 95.67 |
| 90% | 90.67 | 99.67 | 98 | 100 | 100 | 99.67 | 97 | 100 | 98 | 96 |
| 100% | 91 | 100 | 98.33 | 100 | 100 | 99.67 | 97.33 | 100 | 98 | 96.33 |

HMM (GEES)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Speaker  Whi\_Spe subset | 151 | 152 | 153 | 154 | 155 | 156 |
| 10% | 90 | 93.33 | 100 | 100 | 83.33 | 100 |
| 20% | 100 | 96.67 | 100 | 100 | 93.33 | 100 |
| 30% | 100 | 96.67 | 100 | 100 | 90 | 100 |
| 40% | 96.67 | 96.67 | 100 | 100 | 90 | 100 |
| 50% | 100 | 96.67 | 100 | 100 | 90 | 100 |
| 60% | 100 | 96.67 | 100 | 100 | 93.33 | 100 |
| 70% | 100 | 96.67 | 100 | 100 | 93.33 | 100 |
| 80% | 100 | 96.67 | 100 | 100 | 93.33 | 100 |
| 90% | 100 | 96.67 | 100 | 100 | 93.33 | 100 |
| 100% | 100 | 96.67 | 100 | 100 | 93.33 | 100 |

HMM (DB1)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Speaker  Whi\_Spe subset | 101 | 102 | 103 | 104 | 105 | 106 | 109 | 110 | 111 | 112 |
| 10% | 89 | 45 | 92 | 88 | 61 | 54 | 92 | 83 | 89 | 82 |
| 20% | 91 | 53 | 91 | 94 | 70 | 63 | 95 | 89 | 96 | 89 |
| 30% | 92 | 62 | 94 | 95 | 71 | 66 | 92 | 89 | 96 | 90 |
| 40% | 91 | 66 | 95 | 93 | 75 | 68 | 94 | 90 | 96 | 88 |
| 50% | 91 | 64 | 95 | 93 | 76 | 69 | 95 | 92 | 96 | 91 |
| 60% | 90 | 62 | 96 | 94 | 76 | 69 | 95 | 89 | 92 | 90 |
| 70% | 90 | 66 | 96 | 91 | 77 | 69 | 95 | 89 | 92 | 90 |
| 80% | 91 | 63 | 96 | 94 | 77 | 71 | 94 | 89 | 94 | 89 |
| 90% | 91 | 63 | 97 | 90 | 74 | 73 | 94 | 88 | 94 | 89 |
| 100% | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 |

fig3.emf

Figure 3. Average recognition accuracy (in %) for three speech databases in SVM framework. The percent of the Whi-Spe subset used in training is given in abscissa.

SVM (Whi-Spe)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Speaker  Whi\_Spe subset | 2 | 6 | 7 | 8 | 9 | 14 | 15 | 16 | 18 | 19 |
| 10% | 95.59 | 85.97 | 86.77 | 96.99 | 91.78 | 94.38 | 92.38 | 91.58 | 87.17 | 88.78 |
| 20% | 98 | 92.38 | 89.78 | 96.79 | 95.19 | 94.79 | 94.59 | 97.59 | 87.58 | 91.18 |
| 30% | 97.19 | 89.18 | 89.58 | 96.99 | 95.39 | 95.19 | 94.59 | 97.39 | 90.38 | 93.19 |
| 40% | 97.6 | 88.98 | 89.98 | 97.4 | 95.79 | 96.39 | 97.39 | 98.2 | 92.59 | 92.18 |
| 50% | 97.99 | 93.78 | 90.78 | 96.99 | 96.79 | 95.39 | 97.39 | 97.79 | 91.58 | 93.19 |
| 60% | 98.4 | 95.39 | 90.98 | 96.99 | 97.19 | 97.19 | 96.59 | 97.99 | 92.59 | 95.59 |
| 70% | 98.6 | 95.59 | 90.98 | 97.19 | 97.19 | 97.19 | 97.19 | 98.2 | 93.19 | 94.19 |
| 80% | 98.4 | 95.39 | 91.78 | 97.39 | 97.59 | 97.79 | 97.59 | 98.2 | 93.99 | 96.39 |
| 90% | 98.6 | 95.6 | 91.58 | 97.6 | 97.6 | 97.39 | 97.99 | 98.6 | 94.39 | 95.59 |
| 100% | 98.8 | 95.19 | 91.78 | 97.39 | 97.19 | 97.39 | 97.79 | 98.6 | 94.39 | 95.39 |

SVM (GEES)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Speaker  Whi\_Spe subset | 151 | 152 | 153 | 154 | 155 | 156 |
| 10% | 92.59 | 85.18 | 92.59 | 85.18 | 77.78 | 66.67 |
| 20% | 85.18 | 85.18 | 92.59 | 92.59 | 77.78 | 70.37 |
| 30% | 92.59 | 85.18 | 96.3 | 88.89 | 81.48 | 77.78 |
| 40% | 88.88 | 88.89 | 96.3 | 88.89 | 81.48 | 85.18 |
| 50% | 96.3 | 85.18 | 96.3 | 96.3 | 77.78 | 88.89 |
| 60% | 92.59 | 81.48 | 96.3 | 85.18 | 77.78 | 92.59 |
| 70% | 88.88 | 92.59 | 96.3 | 81.48 | 85.18 | 92.59 |
| 80% | 92.59 | 92.59 | 100 | 88.88 | 85.18 | 85.18 |
| 90% | 88.88 | 92.59 | 100 | 81.48 | 92.59 | 81.48 |
| 100% | 92.59 | 96.3 | 96.3 | 100 | 92.59 | 85.1 |

SVM (DB1)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Speaker  Whi\_Spe subset | 101 | 102 | 103 | 104 | 105 | 106 | 109 | 110 | 111 | 112 |
| 10% | 56.56 | 70.7 | 76.77 | 90.9 | 84.84 | 57.58 | 88.89 | 89.89 | 77.78 | 83.84 |
| 20% | 64.64 | 77.78 | 74.75 | 90.9 | 87.88 | 60.6 | 90.9 | 95.96 | 73.74 | 90.9 |
| 30% | 61.61 | 79.8 | 83.84 | 92.93 | 83.83 | 64.65 | 91.92 | 95.96 | 75.75 | 91.92 |
| 40% | 69.69 | 76.76 | 82.83 | 93.93 | 86.86 | 65.66 | 89.9 | 94.95 | 78.79 | 89.9 |
| 50% | 63.63 | 78.79 | 82.83 | 94.95 | 86.86 | 66.67 | 91.92 | 95.96 | 83.83 | 91.92 |
| 60% | 71.71 | 78.79 | 82.83 | 92.93 | 88.88 | 65.66 | 92.93 | 95.96 | 74.75 | 92.93 |
| 70% | 65.66 | 77.79 | 85.86 | 94.95 | 87.88 | 69.7 | 89.9 | 95.96 | 75.76 | 89.9 |
| 80% | 69.69 | 79.8 | 84.85 | 92.93 | 88.89 | 67.68 | 92.93 | 95.96 | 81.82 | 92.93 |
| 90% | 71.71 | 79.8 | 84.85 | 92.93 | 90.91 | 64.65 | 91.92 | 95.96 | 78.78 | 91.92 |
| 100% | 68.68 | 79.8 | 85.86 | 92.93 | 90.91 | 64.65 | 91.92 | 96.97 | 79.8 | 91.92 |

fig4.emf

Figure 4. Average recognition accuracy (in %) for three speech databases in CNN framework. The percent of the Whi-Spe subset used in training is given in abscissa.

CNN (Whi-Spe)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Speaker  Whi\_Spe subset | 2 | 6 | 7 | 8 | 9 | 14 | 15 | 16 | 18 | 19 |
| 10% | 80.6 | 66.2 | 75.4 | 87 | 83 | 85 | 84.6 | 88.6 | 77 | 77 |
| 20% | 88 | 74.6 | 80.8 | 92.8 | 86.8 | 89.2 | 87.2 | 88.8 | 80.8 | 79.2 |
| 30% | 93 | 74.2 | 85.4 | 94.4 | 92.6 | 91.8 | 91.2 | 93 | 86.8 | 88.2 |
| 40% | 94.6 | 84.6 | 86.4 | 97.6 | 93.4 | 95 | 92.8 | 93.8 | 83.4 | 84.8 |
| 50% | 94.2 | 86.6 | 86.6 | 96.6 | 93.2 | 90.6 | 92.6 | 94.8 | 84.6 | 85.2 |
| 60% | 95.8 | 89.4 | 90.4 | 97 | 92.8 | 95.2 | 95 | 92.4 | 88.6 | 85.6 |
| 70% | 95.8 | 84.8 | 86.8 | 97.6 | 94.2 | 94.6 | 94.8 | 94.6 | 91.4 | 91.4 |
| 80% | 95.8 | 89 | 87 | 97.8 | 92.8 | 95.8 | 91.8 | 95 | 91 | 88.4 |
| 90% | 96.2 | 89.2 | 86 | 96.2 | 95.6 | 96.4 | 94.2 | 96 | 91.4 | 89.4 |
| 100% | 94.6 | 88.8 | 87.4 | 97.8 | 96 | 96 | 94.2 | 95.6 | 89.8 | 87.8 |

CNN (GEES)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Speaker  Whi\_Spe subset | 151 | 152 | 153 | 154 | 155 | 156 |
| 10% | 0.6333 | 0.2667 | 0.6 | 0.6 | 0.4 | 0.7 |
| 20% | 0.6333 | 0.5333 | 0.7333 | 0.8667 | 0.4333 | 0.8 |
| 30% | 0.8667 | 0.4667 | 0.7667 | 0.8667 | 0.5333 | 0.8 |
| 40% | 0.8667 | 0.5 | 0.8 | 0.9 | 0.5333 | 0.7 |
| 50% | 0.8667 | 0.6 | 0.8333 | 0.9 | 0.5333 | 0.7667 |
| 60% | 0.8333 | 0.4667 | 0.8 | 0.9 | 0.4667 | 0.8 |
| 70% | 0.9 | 0.4667 | 0.8334 | 0.8667 | 0.6333 | 0.7333 |
| 80% | 0.9 | 0.4333 | 0.8333 | 0.8 | 0.4667 | 0.8 |
| 90% | 0.9333 | 0.7 | 0.7667 | 0.9 | 0.4667 | 0.7667 |
| 100% | 0.9 | 0.5 | 0.8 | 0.93 | 0.47 | 0.87 |

CNN (DB1)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Speaker  Whi\_Spe subset | 101 | 102 | 103 | 104 | 105 | 106 | 109 | 110 | 111 | 112 |
| 10% | 48 | 54 | 56 | 77 | 73 | 56 | 75 | 72 | 63 | 65 |
| 20% | 59 | 78 | 73 | 85 | 70 | 55 | 84 | 82 | 78 | 77 |
| 30% | 67 | 74 | 71 | 82 | 84 | 59 | 77 | 86 | 69 | 80 |
| 40% | 61 | 70 | 70 | 87 | 78 | 60 | 78 | 93 | 71 | 79 |
| 50% | 61 | 78 | 72 | 87 | 81 | 55 | 81 | 92 | 75 | 83 |
| 60% | 71 | 72 | 71 | 84 | 80 | 65 | 84 | 86 | 74 | 87 |
| 70% | 65 | 73 | 77 | 88 | 83 | 62 | 80 | 91 | 75 | 88 |
| 80% | 70 | 76 | 81 | 87 | 82 | 71 | 84 | 93 | 77 | 81 |
| 90% | 64 | 79 | 75 | 91 | 88 | 70 | 90 | 91 | 80 | 83 |
| 100% | 70 | 73 | 84 | 87 | 83 | 75 | 88 | 92 | 79 | 85 |

slika1.emf

Figure 5. Average recognition accuracy (DB1 database) for 7 data augmentation techniques and HMM (a), SVM (b) and CNN (c) framework (Augmented utterances obtained with: P - Python, M - MATLAB. O denotes original training samples). Horizontal dashed line depicts the baseline recognition accuracy using the original Whi-Spe database in training (no augmentation). (p < 0.05 \*; p < 0.005 \*\*; Confidence interval = 95%)

HMM (PS augmentation, Fig. 5a)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| AUG (PYTHON) | 87 | 49 | 89 | 93 | 76 | 66 | 91 | 86 | 88 | 93 | **81,8** |
| AUG (MATLAB) | 95 | 73 | 97 | 96 | 79 | 77 | 96 | 90 | 92 | 91 | **88,6** |
| FUSION (ORIG+PYT) | 91 | 64 | 95 | 95 | 78 | 73 | 94 | 91 | 92 | 91 | **86,4** |
| FUSION (ORIG+MAT) | 94 | 68 | 96 | 95 | 80 | 75 | 94 | 89 | 92 | 90 | **87,3** |
| FUSION (ORIG+PYT+MAT) | 92 | 67 | 94 | 96 | 78 | 73 | 95 | 91 | 92 | 91 | **86,9** |

HMM (TST augmentation, Fig. 5a)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| AUG (PYTHON) | 89 | 46 | 84 | 92 | 68 | 61 | 94 | 73 | 86 | 90 | **78,3** |
| AUG (MATLAB) | 94 | 64 | 95 | 98 | 79 | 73 | 97 | 87 | 88 | 92 | **86,7** |
| FUSION (ORIG+PYT) | 92 | 63 | 92 | 97 | 77 | 73 | 95 | 89 | 93 | 91 | **86,2** |
| FUSION (ORIG+MAT) | 93 | 67 | 99 | 98 | 78 | 74 | 94 | 93 | 90 | 91 | **87,7** |
| FUSION (ORIG+PYT+MAT) | 92 | 61 | 98 | 99 | 78 | 74 | 96 | 91 | 90 | 91 | **87** |

HMM (VC augmentation, Fig. 5a)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| AUG (PYTHON) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| AUG (MATLAB) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| FUSION (ORIG+PYT) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| FUSION (ORIG+MAT) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| FUSION (ORIG+PYT+MAT) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |

HMM (PS+TST augmentation, Fig. 5a)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| AUG (PYTHON) | 89 | 54 | 90 | 89 | 75 | 71 | 92 | 84 | 90 | 90 | **82,4** |
| AUG (MATLAB) | 93 | 66 | 97 | 98 | 80 | 75 | 94 | 91 | 91 | 91 | **87,6** |
| FUSION (ORIG+PYT) | 92 | 67 | 95 | 97 | 78 | 72 | 95 | 90 | 92 | 91 | **86,9** |
| FUSION (ORIG+MAT) | 93 | 67 | 97 | 96 | 79 | 74 | 93 | 89 | 92 | 90 | **87** |
| FUSION (ORIG+PYT+MAT) | 92 | 66 | 98 | 97 | 79 | 73 | 95 | 91 | 92 | 92 | **87,5** |

HMM (PS+VC augmentation, Fig. 5a)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| AUG (PYTHON) | 90 | 68 | 96 | 95 | 75 | 71 | 95 | 91 | 92 | 91 | **86,4** |
| AUG (MATLAB) | 92 | 65 | 96 | 96 | 79 | 75 | 92 | 88 | 93 | 90 | **86,6** |
| FUSION (ORIG+PYT) | 92 | 70 | 97 | 95 | 73 | 71 | 94 | 91 | 92 | 89 | **86,4** |
| FUSION (ORIG+MAT) | 92 | 65 | 98 | 95 | 80 | 74 | 94 | 89 | 94 | 90 | **87,1** |
| FUSION (ORIG+PYT+MAT) | 93 | 68 | 97 | 95 | 78 | 73 | 94 | 91 | 92 | 90 | **87,1** |

HMM (TST+VC augmentation, Fig. 5a)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| AUG (PYTHON) | 93 | 60 | 92 | 98 | 77 | 73 | 94 | 89 | 92 | 90 | **85,8** |
| AUG (MATLAB) | 92 | 64 | 98 | 97 | 81 | 73 | 94 | 89 | 90 | 91 | **86,9** |
| FUSION (ORIG+PYT) | 92 | 68 | 98 | 97 | 79 | 71 | 93 | 88 | 92 | 90 | **86,8** |
| FUSION (ORIG+MAT) | 92 | 65 | 98 | 95 | 77 | 74 | 94 | 89 | 90 | 90 | **86,4** |
| FUSION (ORIG+PYT+MAT) | 92 | 67 | 98 | 96 | 80 | 72 | 93 | 89 | 92 | 90 | **86,9** |

HMM (PS+TST+VC augmentation, Fig. 5a)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| AUG (PYTHON) | 90 | 51 | 90 | 89 | 71 | 65 | 92 | 82 | 90 | 89 | **80,9** |
| AUG (MATLAB) | 94 | 66 | 95 | 98 | 80 | 73 | 95 | 91 | 90 | 92 | **87,4** |
| FUSION (ORIG+PYT) | 92 | 68 | 96 | 97 | 77 | 69 | 94 | 90 | 92 | 92 | **86,7** |
| FUSION (ORIG+MAT) | 94 | 68 | 99 | 95 | 77 | 73 | 94 | 89 | 92 | 91 | **87,2** |
| FUSION (ORIG+PYT+MAT) | 92 | 68 | 99 | 99 | 80 | 72 | 95 | 89 | 92 | 92 | **87,8** |

SVM (PS augmentation, Fig. 5b)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 68,68 | 79,8 | 85,86 | 92,93 | 90,91 | 64,65 | 91,92 | 96,97 | 79,8 | 91,92 | **84,344** |
| AUG (PYTHON) | 70,71 | 78,79 | 85,86 | 92,93 | 90,91 | 76,77 | 87,88 | 95,96 | 82,83 | 85,86 | **84,85** |
| AUG (MATLAB) | 68,69 | 79,8 | 85,86 | 93,94 | 90,91 | 73,74 | 88,89 | 95,96 | 83,84 | 89,9 | **85,153** |
| FUSION (ORIG+PYT) | 70,71 | 80,81 | 85,86 | 92,93 | 90,91 | 75,76 | 88,89 | 95,96 | 82,83 | 86,87 | **85,153** |
| FUSION (ORIG+MAT) | 69,7 | 79,79 | 84,85 | 91,92 | 90,91 | 71,72 | 88,89 | 95,96 | 82,83 | 89,89 | **84,646** |
| FUSION (ORIG+PYT+MAT) | 69,7 | 79,8 | 86,87 | 91,92 | 90,91 | 76,77 | 87,88 | 95,96 | 83,84 | 88,89 | **85,254** |

SVM (TST augmentation, Fig. 5b)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 68,68 | 79,8 | 85,86 | 92,93 | 90,91 | 64,65 | 91,92 | 96,97 | 79,8 | 91,92 | **84,344** |
| AUG (PYTHON) | 71,72 | 76,77 | 82,83 | 91,92 | 89,89 | 68,69 | 90,91 | 95,96 | 79,8 | 89,9 | **83,839** |
| AUG (MATLAB) | 65,66 | 76,77 | 83,84 | 92,93 | 87,88 | 67,68 | 89,9 | 95,96 | 78,79 | 85,86 | **82,527** |
| FUSION (ORIG+PYT) | 69,7 | 80,81 | 86,87 | 92,93 | 88,89 | 66,67 | 83,84 | 96,97 | 80,81 | 86,87 | **83,436** |
| FUSION (ORIG+MAT) | 67,68 | 80,81 | 83,84 | 95,96 | 86,87 | 65,66 | 85,86 | 95,96 | 79,8 | 87,88 | **83,032** |
| FUSION (ORIG+PYT+MAT) | 69,7 | 77,78 | 84,85 | 92,93 | 90,91 | 68,68 | 91,92 | 96,97 | 79,8 | 91,92 | **84,546** |

SVM (VC augmentation, Fig. 5b)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 68,68 | 79,8 | 85,86 | 92,93 | 90,91 | 64,65 | 91,92 | 96,97 | 79,8 | 91,92 | **84,344** |
| AUG (PYTHON) | 69,7 | 79,8 | 84,85 | 92,93 | 90,91 | 65,66 | 91,92 | 96,97 | 80,81 | 87,88 | **84,143** |
| AUG (MATLAB) | 68,69 | 79,8 | 84,85 | 92,93 | 90,91 | 64,65 | 92,93 | 95,96 | 80,81 | 88,89 | **84,042** |
| FUSION (ORIG+PYT) | 67,68 | 80,81 | 84,85 | 95,96 | 86,87 | 66,67 | 84,85 | 96,97 | 77,78 | 86,87 | **82,931** |
| FUSION (ORIG+MAT) | 67,68 | 80,81 | 84,85 | 95,96 | 87,88 | 65,66 | 84,85 | 96,97 | 76,77 | 85,86 | **82,729** |
| FUSION (ORIG+PYT+MAT) | 69,7 | 80,81 | 84,85 | 92,93 | 90,91 | 64,65 | 92,93 | 95,96 | 78,79 | 87,88 | **83,941** |

SVM (PS+TST augmentation, Fig. 5b)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 68,68 | 79,8 | 85,86 | 92,93 | 90,91 | 64,65 | 91,92 | 96,97 | 79,8 | 91,92 | **84,344** |
| AUG (PYTHON) | 69,7 | 81,82 | 85,86 | 90,91 | 88,89 | 72,73 | 87,88 | 95,96 | 82,83 | 87,88 | **84,446** |
| AUG (MATLAB) | 66,67 | 79,8 | 86,87 | 92,93 | 89,89 | 72,73 | 90,91 | 95,96 | 81,82 | 88,89 | **84,647** |
| FUSION (ORIG+PYT) | 69,7 | 79,8 | 86,87 | 91,92 | 88,89 | 67,68 | 83,84 | 95,96 | 82,83 | 82,82 | **83,031** |
| FUSION (ORIG+MAT) | 66,67 | 81,82 | 85,86 | 92,93 | 89,9 | 71,72 | 89,9 | 95,96 | 79,8 | 86,87 | **84,143** |
| FUSION (ORIG+PYT+MAT) | 68,69 | 80,81 | 85,86 | 91,92 | 90,91 | 74,75 | 88,89 | 95,96 | 83,84 | 88,89 | **85,052** |

SVM (PS+VC augmentation, Fig. 5b)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 68,68 | 79,8 | 85,86 | 92,93 | 90,91 | 64,65 | 91,92 | 96,97 | 79,8 | 91,92 | **84,344** |
| AUG (PYTHON) | 68,69 | 82,83 | 85,86 | 93,94 | 91,92 | 71,72 | 88,89 | 94,95 | 82,83 | 87,88 | **84,951** |
| AUG (MATLAB) | 68,69 | 77,78 | 83,84 | 93,94 | 90,91 | 72,73 | 89,9 | 95,96 | 82,83 | 89,9 | **84,648** |
| FUSION (ORIG+PYT) | 69,7 | 80,81 | 86,87 | 93,94 | 86,87 | 68,69 | 84,85 | 95,96 | 80,81 | 84,85 | **83,335** |
| FUSION (ORIG+MAT) | 66,67 | 79,79 | 84,85 | 94,95 | 84,85 | 66,67 | 86,87 | 96,97 | 79,8 | 85,86 | **82,728** |
| FUSION (ORIG+PYT+MAT) | 70,71 | 81,82 | 85,86 | 92,93 | 91,92 | 69,7 | 89,9 | 95,96 | 82,83 | 86,87 | **84,85** |

SVM (TST+VC augmentation, Fig. 5b)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 68,68 | 79,8 | 85,86 | 92,93 | 90,91 | 64,65 | 91,92 | 96,97 | 79,8 | 91,92 | **84,344** |
| AUG (PYTHON) | 70,71 | 79,8 | 82,83 | 93,94 | 90,91 | 67,68 | 90,91 | 95,96 | 79,8 | 85,86 | **83,84** |
| AUG (MATLAB) | 72,45 | 78,79 | 84,85 | 93,94 | 89,9 | 67,68 | 90,91 | 94,95 | 77,78 | 88,89 | **84,014** |
| FUSION (ORIG+PYT) | 69,7 | 84,85 | 86,87 | 94,95 | 87,88 | 65,66 | 85,86 | 96,97 | 79,8 | 87,88 | **84,042** |
| FUSION (ORIG+MAT) | 71,72 | 79,8 | 85,86 | 95,96 | 87,88 | 62,63 | 84,85 | 95,96 | 77,78 | 87,88 | **83,032** |
| FUSION (ORIG+PYT+MAT) | 70,71 | 78,79 | 85,86 | 91,92 | 90,91 | 67,68 | 91,92 | 94,95 | 77,78 | 89,9 | **84,042** |

SVM (PS+TST+VC augmentation, Fig. 5b)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 68,68 | 79,8 | 85,86 | 92,93 | 90,91 | 64,65 | 91,92 | 96,97 | 79,8 | 91,92 | **84,344** |
| AUG (PYTHON) | 72,73 | 77,78 | 84,85 | 91,92 | 90,91 | 71,72 | 86,87 | 95,96 | 78,79 | 85,86 | **83,739** |
| AUG (MATLAB) | 68,69 | 79,8 | 83,84 | 91,92 | 90,91 | 71,72 | 90,91 | 95,96 | 82,83 | 86,87 | **84,345** |
| FUSION (ORIG+PYT) | 72,73 | 77,78 | 86,87 | 93,94 | 87,88 | 68,69 | 85,86 | 95,96 | 80,81 | 85,86 | **83,638** |
| FUSION (ORIG+MAT) | 69,7 | 81,82 | 86,87 | 94,95 | 87,88 | 69,7 | 86,87 | 95,96 | 81,82 | 88,89 | **84,446** |
| FUSION (ORIG+PYT+MAT) | 70,71 | 79,8 | 85,86 | 92,93 | 91,92 | 72,73 | 89,9 | 95,96 | 82,83 | 86,87 | **84,951** |

CNN (PS augmentation, Fig. 5c)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 67,3 | 72,7 | 77,3 | 88,4 | 83,8 | 66,0 | 85,7 | 92,5 | 79,1 | 83,2 | **79,60** |
| AUG (PYTHON) | 67,8 | 76,2 | 75,6 | 85,1 | 86,7 | 71,4 | 83,9 | 93,3 | 81,8 | 84,1 | **80.59** |
| AUG (MATLAB) | 68.0 | 73.9 | 78.0 | 84.0 | 83.8 | 68.8 | 83.9 | 93.3 | 83.0 | 83.6 | **80.03** |
| FUSION (ORIG+PYT) | 68.3 | 78.1 | 81.8 | 88.0 | 87.7 | 75.3 | 86.6 | 94.9 | 82.2 | 85.1 | **82.80** |
| FUSION (ORIG+MAT) | 67.4 | 74.0 | 81.8 | 88.4 | 86.1 | 72.6 | 87.5 | 93.9 | 81.0 | 85.5 | **81.82** |
| FUSION (ORIG+PYT+MAT) | 68,7 | 77,8 | 81,0 | 86,9 | 87,9 | 75,3 | 87,3 | 95,6 | 83,7 | 85,1 | **82,93** |

CNN (TST augmentation, Fig. 5c)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 67,3 | 72,7 | 77,3 | 88,4 | 83,8 | 66,0 | 85,7 | 92,5 | 79,1 | 83,2 | **79,60** |
| AUG (PYTHON) | 67.6 | 75.4 | 76.6 | 87.5 | 85.6 | 69.5 | 84.5 | 92.7 | 81.4 | 85.7 | **80,65** |
| AUG (MATLAB) | 68.1 | 72.7 | 77.1 | 87.7 | 83.4 | 67.7 | 86.9 | 91.9 | 80.5 | 84.9 | **80.09** |
| FUSION (ORIG+PYT) | 68.6 | 75.2 | 80.3 | 89.5 | 84.5 | 72.6 | 87.9 | 93.8 | 82.2 | 86.2 | **82.08** |
| FUSION (ORIG+MAT) | 66.8 | 74.3 | 80.2 | 89.5 | 85.5 | 73.7 | 87.2 | 92.2 | 80.2 | 85.5 | **81.51** |
| FUSION (ORIG+PYT+MAT) | 69.5 | 74.7 | 81.3 | 88.4 | 86.6 | 73.9 | 87.6 | 93.9 | 82.2 | 86.7 | **82.48** |

CNN (VC augmentation, Fig. 5c)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 67,3 | 72,7 | 77,3 | 88,4 | 83,8 | 66,0 | 85,7 | 92,5 | 79,1 | 83,2 | **79,60** |
| AUG (PYTHON) | 65.5 | 73.3 | 77.7 | 86.9 | 85.3 | 69.9 | 85.4 | 92.1 | 80.7 | 83.6 | **80.04** |
| AUG (MATLAB) | 66.6 | 72.3 | 75.6 | 88.2 | 84.3 | 67.8 | 86.3 | 92.1 | 79.7 | 84.9 | **79.78** |
| FUSION (ORIG+PYT) | 67.4 | 75.4 | 78.2 | 88.0 | 85.3 | 73.9 | 86.5 | 93.5 | 80.8 | 83.5 | **81.25** |
| FUSION (ORIG+MAT) | 66.6 | 74.4 | 81.1 | 88.6 | 86.1 | 72.1 | 87.8 | 92.4 | 82.4 | 83.5 | **81.50** |
| FUSION (ORIG+PYT+MAT) | 66.9 | 73.1 | 79.8 | 88.7 | 84.2 | 71.4 | 87.2 | 94.0 | 80.4 | 85.0 | **81,07** |

CNN (PS+TST augmentation, Fig. 5c)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 67,3 | 72,7 | 77,3 | 88,4 | 83,8 | 66,0 | 85,7 | 92,5 | 79,1 | 83,2 | **79,60** |
| AUG (PYTHON) | 68.9 | 78.5 | 77.2 | 87 | 85.7 | 72.8 | 85.4 | 92.8 | 79 | 82.6 | **80.99** |
| AUG (MATLAB) | 66.8 | 74.4 | 76 | 86.6 | 84.8 | 71.2 | 85.2 | 93.1 | 79.9 | 84.7 | **80.27** |
| FUSION (ORIG+PYT) | 71.2 | 76.5 | 80.9 | 88 | 87.1 | 75.9 | 86.6 | 94.5 | 80.8 | 83.4 | **82.49** |
| FUSION (ORIG+MAT) | 68.1 | 75.2 | 79.7 | 88.3 | 86.1 | 73.6 | 87.2 | 94.3 | 81 | 84.9 | **81,84** |
| FUSION (ORIG+PYT+MAT) | 70.4 | 75.3 | 81.9 | 88.3 | 85.6 | 77.6 | 89.1 | 94.4 | 81.7 | 84.8 | **82,91** |

CNN (PS+VC augmentation, Fig. 5c)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 67,3 | 72,7 | 77,3 | 88,4 | 83,8 | 66,0 | 85,7 | 92,5 | 79,1 | 83,2 | **79,60** |
| AUG (PYTHON) | 65,6 | 75,2 | 78 | 87,9 | 85 | 73,4 | 85,7 | 93,8 | 82,7 | 83,3 | **81,06** |
| AUG (MATLAB) | 66,4 | 73,1 | 78,7 | 86,8 | 86,7 | 69,7 | 87 | 93 | 79,7 | 84,3 | **80,54** |
| FUSION (ORIG+PYT) | 67.4 | 75.5 | 80.6 | 88.0 | 86.1 | 75.4 | 86.6 | 94.3 | 81.2 | 85.6 | **82.07** |
| FUSION (ORIG+MAT) | 68.3 | 75.6 | 78.9 | 89.0 | 87.2 | 75.8 | 86.5 | 93.9 | 80.4 | 84.0 | **81.96** |
| FUSION (ORIG+PYT+MAT) | 69.1 | 74.9 | 80.2 | 88.7 | 87.8 | 75.1 | 87.9 | 94.4 | 82.3 | 86.2 | **82.66** |

CNN (TST+VC augmentation, Fig. 5c)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 67,3 | 72,7 | 77,3 | 88,4 | 83,8 | 66,0 | 85,7 | 92,5 | 79,1 | 83,2 | **79,60** |
| AUG (PYTHON) | 67,7 | 76 | 76,6 | 87,4 | 83,6 | 70,4 | 86,2 | 94,3 | 81,7 | 85 | **80,89** |
| AUG (MATLAB) | 66,3 | 72,1 | 76,7 | 85,6 | 82,8 | 69,4 | 86,9 | 92,7 | 77,9 | 84,3 | **79,47** |
| FUSION (ORIG+PYT) | 68.3 | 75.3 | 79.2 | 88.6 | 86.2 | 70.0 | 87.3 | 92.2 | 81.1 | 83.9 | **81.21** |
| FUSION (ORIG+MAT) | 67.8 | 75.9 | 80.9 | 88.9 | 85.6 | 73.0 | 89.4 | 93.7 | 79.3 | 84.4 | **81.89** |
| FUSION (ORIG+PYT+MAT) | 68.2 | 76.1 | 80.3 | 88.8 | 86.1 | 73.4 | 89 | 93.78 | 81.8 | 85.8 | **82.33** |

CNN (PS+TST+VC augmentation, Fig. 5c)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 67,3 | 72,7 | 77,3 | 88,4 | 83,8 | 66,0 | 85,7 | 92,5 | 79,1 | 83,2 | **79,60** |
| AUG (PYTHON) | 68,3 | 76,4 | 76,7 | 87,1 | 84,5 | 73,8 | 83,8 | 93,7 | 80,8 | 83,2 | **80.83** |
| AUG (MATLAB) | 67,1 | 75,2 | 78,7 | 85,9 | 87,4 | 72 | 85,5 | 93 | 80,7 | 83,4 | **80.89** |
| FUSION (ORIG+PYT) | 69 | 75.1 | 80.2 | 86.4 | 87.3 | 74 | 86.3 | 94.5 | 80.7 | 84.1 | **81.76** |
| FUSION (ORIG+MAT) | 69.1 | 74.6 | 80.8 | 88.3 | 87.5 | 72.6 | 87.9 | 93.1 | 80.3 | 84.9 | **81,91** |
| FUSION (ORIG+PYT+MAT) | 67.3 | 74.5 | 82.9 | 87.5 | 86.4 | 74.7 | 85.5 | 95.6 | 81.9 | 85.9 | **82.22** |

slika2

Figure 6. Average recognition accuracy (in %) for DB1 database in HMM (a), SVM (b) and CNN (c) framework (Augmented utterances obtained with: P - Python, M - MATLAB. O denotes original training samples). Horizontal dashed lines depict the recognition without augmentation (long - full capacity; short - corresponding Whi-Spe subset percentage).

HMM (PS augmentation, Fig. 6a) – 10%, PS augmentation, number of augmentations 10

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 89 | 45 | 92 | 88 | 61 | 54 | 92 | 83 | 89 | 82 | **77,5** |
| NO AUG Full (ORIGINAL) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| AUG (PYTHON) | 88 | 44 | 90 | 85 | 68 | 65 | 92 | 77 | 89 | 87 | **78,5** |
| AUG (MATLAB) | 93 | 56 | 94 | 95 | 75 | 70 | 91 | 90 | 92 | 89 | **84,5** |
| FUSION (ORIG+PYT) | 87 | 49 | 92 | 86 | 71 | 67 | 93 | 81 | 89 | 88 | **80,3** |
| FUSION (ORIG+MAT) | 91 | 63 | 91 | 94 | 74 | 66 | 92 | 84 | 91 | 88 | **83,4** |
| FUSION (ORIG+PYT+MAT) | 93 | 55 | 95 | 95 | 75 | 69 | 92 | 88 | 92 | 88 | **84,2** |

HMM (PS augmentation, Fig. 6a) – 20%, PS augmentation, number of augmentations 5

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 91 | 53 | 91 | 94 | 70 | 63 | 95 | 89 | 96 | 89 | **83,1** |
| NO AUG Full (ORIGINAL) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| AUG (PYTHON) | 89 | 53 | 89 | 91 | 75 | 65 | 95 | 81 | 88 | 86 | **81,2** |
| AUG (MATLAB) | 95 | 61 | 96 | 98 | 77 | 70 | 94 | 90 | 93 | 89 | **86,3** |
| FUSION (ORIG+PYT) | 89 | 61 | 91 | 95 | 78 | 66 | 93 | 87 | 91 | 91 | **84,2** |
| FUSION (ORIG+MAT) | 93 | 64 | 95 | 99 | 77 | 66 | 94 | 89 | 93 | 89 | **85,9** |
| FUSION (ORIG+PYT+MAT) | 94 | 62 | 95 | 99 | 79 | 67 | 96 | 88 | 91 | 90 | **86,1** |

HMM (PS augmentation, Fig. 6a) – 50%, PS augmentation, number of augmentations 2

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 91 | 64 | 95 | 93 | 76 | 69 | 95 | 92 | 96 | 91 | **86,2** |
| NO AUG Full (ORIGINAL) | 91 | 63 | 98 | 93 | 75 | 70 | 95 | 88 | 94 | 89 | **85,6** |
| AUG (PYTHON) | 87 | 51 | 91 | 93 | 69 | 63 | 94 | 81 | 87 | 90 | **80,6** |
| AUG (MATLAB) | 94 | 66 | 96 | 98 | 77 | 75 | 93 | 91 | 92 | 90 | **87,2** |
| FUSION (ORIG+PYT) | 90 | 64 | 96 | 95 | 77 | 70 | 95 | 85 | 92 | 91 | **85,5** |
| FUSION (ORIG+MAT) | 95 | 70 | 95 | 97 | 80 | 75 | 94 | 90 | 94 | 92 | **88,2** |
| FUSION (ORIG+PYT+MAT) | 94 | 65 | 96 | 98 | 78 | 71 | 96 | 89 | 94 | 90 | **87,1** |

SVM (PS augmentation, Fig. 6b) – 10%, PS augmentation, number of augmentations 10

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 56,56 | 70,7 | 76,77 | 90,9 | 84,84 | 57,58 | 88,89 | 89,89 | 77,78 | 83,84 | 77,775 |
| NO AUG Full (ORIGINAL) | 68,68 | 79,8 | 85,86 | 92,93 | 90,91 | 64,65 | 91,92 | 96,97 | 79,8 | 91,92 | 84,344 |
| AUG (PYTHON) | 72,73 | 80,81 | 77,78 | 88,89 | 88,89 | 67,67 | 84,85 | 94,95 | 81,82 | 85,86 | 82,425 |
| AUG (MATLAB) | 70,71 | 76,77 | 75,76 | 88,89 | 82,83 | 67,68 | 85,86 | 93,94 | 86,87 | 83,84 | 81,315 |
| FUSION (ORIG+PYT) | 71,71 | 80,81 | 76,77 | 88,89 | 87,88 | 68,69 | 84,85 | 93,94 | 81,82 | 85,86 | 82,122 |
| FUSION (ORIG+MAT) | 70,71 | 77,78 | 76,77 | 87,88 | 82,83 | 66,67 | 85,86 | 93,94 | 85,86 | 84,85 | 81,315 |
| FUSION (ORIG+PYT+MAT) | 69,69697 | 80,80808 | 74,74747 | 88,88889 | 85,85859 | 67,67677 | 85,85859 | 93,93939 | 83,83838 | 84,84848 | 81,616161 |

SVM (PS augmentation, Fig. 6b) – 20%, PS augmentation, number of augmentations 5

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 64,64 | 77,78 | 74,75 | 90,9 | 87,88 | 60,6 | 90,9 | 95,96 | 73,74 | 90,9 | 80,805 |
| NO AUG Full (ORIGINAL) | 68,68 | 79,8 | 85,86 | 92,93 | 90,91 | 64,65 | 91,92 | 96,97 | 79,8 | 91,92 | 84,344 |
| AUG (PYTHON) | 66,67 | 81,82 | 74,75 | 89,9 | 85,86 | 65,66 | 85,86 | 95,96 | 85,86 | 80,81 | 81,315 |
| AUG (MATLAB) | 66,67 | 74,75 | 74,75 | 89,9 | 82,83 | 63,64 | 84,85 | 94,95 | 85,86 | 83,84 | 80,204 |
| FUSION (ORIG+PYT) | 68,69 | 79,8 | 77,78 | 90,91 | 86,87 | 67,68 | 88,89 | 95,96 | 86,87 | 82,83 | 82,628 |
| FUSION (ORIG+MAT) | 68,69 | 76,77 | 78,79 | 91,92 | 83,84 | 68,69 | 89,9 | 95,96 | 85,86 | 86,87 | 82,729 |
| FUSION (ORIG+PYT+MAT) | 70,70707 | 80,80808 | 81,81818 | 89,89899 | 89,89899 | 71,71717 | 87,87879 | 95,9596 | 86,86869 | 87,87879 | 84,343435 |

SVM (PS augmentation, Fig. 6b) – 50%, PS augmentation, number of augmentations 2

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) | 63,63 | 78,79 | 82,83 | 94,95 | 86,86 | 66,67 | 91,92 | 95,96 | 83,83 | 91,92 | 83,736 |
| NO AUG Full (ORIGINAL) | 68,68 | 79,8 | 85,86 | 92,93 | 90,91 | 64,65 | 91,92 | 96,97 | 79,8 | 91,92 | 84,344 |
| AUG (PYTHON) | 71,72 | 81,82 | 80,81 | 93,94 | 86,87 | 70,71 | 88,88 | 94,95 | 80,81 | 85,86 | 83,637 |
| AUG (MATLAB) | 71,72 | 78,79 | 80,81 | 93,94 | 88,88 | 72,73 | 91,92 | 95,96 | 83,84 | 87,88 | 84,647 |
| FUSION (ORIG+PYT) | 72,73 | 81,82 | 82,83 | 93,94 | 89,9 | 71,72 | 90,91 | 95,96 | 80,81 | 86,87 | 84,749 |
| FUSION (ORIG+MAT) | 69,7 | 80,81 | 81,82 | 92,93 | 88,89 | 73,74 | 91,92 | 95,96 | 84 | 88,89 | 84,866 |
| FUSION (ORIG+PYT+MAT) | 72,72727 | 81,81818 | 83,83838 | 93,93939 | 88,88889 | 71,71717 | 90,90909 | 95,9596 | 80,80808 | 87,87879 | 84,848484 |

CNN (PS augmentation, Fig. 6c) – 10%, PS augmentation, number of augmentations 10

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) |  |  |  |  |  |  |  |  |  |  |  |
| NO AUG Full (ORIGINAL) |  |  |  |  |  |  |  |  |  |  |  |
| AUG (PYTHON) |  |  |  |  |  |  |  |  |  |  |  |
| AUG (MATLAB) |  |  |  |  |  |  |  |  |  |  |  |
| FUSION (ORIG+PYT) |  |  |  |  |  |  |  |  |  |  |  |
| FUSION (ORIG+MAT) |  |  |  |  |  |  |  |  |  |  |  |
| FUSION (ORIG+PYT+MAT) |  |  |  |  |  |  |  |  |  |  |  |

CNN (PS augmentation, Fig. 6c) – 20%, PS augmentation, number of augmentations 5

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) |  |  |  |  |  |  |  |  |  |  |  |
| NO AUG Full (ORIGINAL) |  |  |  |  |  |  |  |  |  |  |  |
| AUG (PYTHON) |  |  |  |  |  |  |  |  |  |  |  |
| AUG (MATLAB) |  |  |  |  |  |  |  |  |  |  |  |
| FUSION (ORIG+PYT) |  |  |  |  |  |  |  |  |  |  |  |
| FUSION (ORIG+MAT) |  |  |  |  |  |  |  |  |  |  |  |
| FUSION (ORIG+PYT+MAT) |  |  |  |  |  |  |  |  |  |  |  |

CNN (PS augmentation, Fig. 6c) – 50%, PS augmentation, number of augmentations 2

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (DB1 database) | | | | | | | | | |  |
| **101** | **102** | **103** | **104** | **105** | **106** | **109** | **110** | **111** | **112** | **AVG** |
| NO AUG (ORIGINAL) |  |  |  |  |  |  |  |  |  |  |  |
| NO AUG Full (ORIGINAL) |  |  |  |  |  |  |  |  |  |  |  |
| AUG (PYTHON) |  |  |  |  |  |  |  |  |  |  |  |
| AUG (MATLAB) |  |  |  |  |  |  |  |  |  |  |  |
| FUSION (ORIG+PYT) |  |  |  |  |  |  |  |  |  |  |  |
| FUSION (ORIG+MAT) |  |  |  |  |  |  |  |  |  |  |  |
| FUSION (ORIG+PYT+MAT) |  |  |  |  |  |  |  |  |  |  |  |

HMM (PS augmentation, Table II) – PS augmentation, number of augmentations 10 (Subset percentage 10%)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (GEES database) | | | | | |  |
| **151** | **152** | **153** | **154** | **155** | **156** | **AVG** |
| NO AUG (ORIGINAL) | 90 | 93.33 | 100 | 100 | 83.33 | 100 | 94.44 |
| AUG (PYTHON) | 86.67 | 76.67 | 90 | 96.67 | 83.33 | 90 | 87.22 |
| AUG (MATLAB) | 96.67 | 93.33 | 100 | 100 | 93.33 | 96.67 | 96.67 |
| FUSION (ORIG+PYT) | 86.67 | 83.33 | 93.33 | 100 | 83.33 | 90 | 89.44 |
| FUSION (ORIG+MAT) | 96.67 | 90 | 100 | 100 | 93.33 | 96.67 | 96.11 |
| FUSION (ORIG+PYT+MAT) | 96.67 | 86.67 | 100 | 100 | 90 | 93.33 | 94.45 |

HMM (PS augmentation, Table II) – PS augmentation, number of augmentations 5 (Subset percentage 20%)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (GEES database) | | | | | |  |
| **151** | **152** | **153** | **154** | **155** | **156** | **AVG** |
| NO AUG (ORIGINAL) | 100 | 96.67 | 100 | 100 | 90 | 100 | 97.78 |
| AUG (PYTHON) | 90 | 83.33 | 93.33 | 100 | 83.33 | 93.33 | 90.55 |
| AUG (MATLAB) | 96.67 | 93.33 | 100 | 100 | 93.33 | 100 | 97.22 |
| FUSION (ORIG+PYT) | 93.33 | 86.67 | 100 | 100 | 90 | 93.33 | 93.89 |
| FUSION (ORIG+MAT) | 96.67 | 96.67 | 100 | 100 | 90 | 100 | 97.22 |
| FUSION (ORIG+PYT+MAT) | 96.67 | 93.33 | 100 | 100 | 90 | 100 | 96.67 |

HMM (PS augmentation, Table II) – PS augmentation, number of augmentations 2 (Subset percentage 50%)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (GEES database) | | | | | |  |
| **151** | **152** | **153** | **154** | **155** | **156** | **AVG** |
| NO AUG (ORIGINAL) | 100 | 96.67 | 100 | 100 | 90 | 100 | 97.78 |
| AUG (PYTHON) | 86.67 | 83.33 | 96.67 | 100 | 86.67 | 90 | 90.55 |
| AUG (MATLAB) | 96.67 | 96.67 | 100 | 100 | 96.67 | 100 | 98.33 |
| FUSION (ORIG+PYT) | 96.67 | 93.33 | 100 | 100 | 90 | 96.67 | 96.11 |
| FUSION (ORIG+MAT) | 100 | 96.67 | 100 | 100 | 93.33 | 100 | 98.33 |
| FUSION (ORIG+PYT+MAT) | 96.67 | 96.67 | 100 | 100 | 90 | 100 | 97.22 |

HMM (PS augmentation, Table II) – PS augmentation, number of augmentations 1 (FULL)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (GEES database) | | | | | |  |
| **151** | **152** | **153** | **154** | **155** | **156** | **AVG** |
| NO AUG (ORIGINAL) | 100 | 96.67 | 100 | 100 | 93.33 | 100 | 98.33 |
| AUG (PYTHON) | 93.33 | 83.33 | 100 | 100 | 86.67 | 96.67 | 93.33 |
| AUG (MATLAB) | 100 | 96.67 | 100 | 100 | 93.33 | 100 | 98.33 |
| FUSION (ORIG+PYT) | 100 | 96.67 | 100 | 100 | 90 | 100 | 97.78 |
| FUSION (ORIG+MAT) | 100 | 96.67 | 100 | 100 | 93.33 | 100 | 98.33 |
| FUSION (ORIG+PYT+MAT) | 100 | 96.67 | 100 | 100 | 90 | 100 | 97.78 |

SVM (PS augmentation, Table III) – PS augmentation, number of augmentations 10 (Subset percentage 10%)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (GEES database) | | | | | |  |
| **151** | **152** | **153** | **154** | **155** | **156** | **AVG** |
| NO AUG (ORIGINAL) | 86.67 | 85.19 | 93.33 | 82.97 | 73.33 | 75.56 | 82.84 |
| AUG (PYTHON) | 87.41 | 80 | 96.30 | 87.41 | 71.11 | 74.07 | 82.72 |
| AUG (MATLAB) | 86.67 | 82.96 | 92.59 | 83.70 | 72.59 | 73.33 | 81.97 |
| FUSION (ORIG+PYT) | 87.41 | 79.26 | 97.04 | 85.93 | 71.11 | 72.59 | 82.22 |
| FUSION (ORIG+MAT) | 86.67 | 81.48 | 93.33 | 87.41 | 73.33 | 77.04 | 83.21 |
| FUSION (ORIG+PYT+MAT) | 85.93 | 79.26 | 94.07 | 87.41 | 70.37 | 77.04 | 82.35 |

SVM (PS augmentation, Table III) – PS augmentation, number of augmentations 5 (Subset percentage 20%)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (GEES database) | | | | | |  |
| **151** | **152** | **153** | **154** | **155** | **156** | **AVG** |
| NO AUG (ORIGINAL) | 89.63 | 88.15 | 95.56 | 85.91 | 77.04 | 75 | 85.21 |
| AUG (PYTHON) | 90.37 | 82.22 | 92.59 | 90.37 | 74.81 | 94.07 | 87.41 |
| AUG (MATLAB) | 93.33 | 86.67 | 97.04 | 87.41 | 77.04 | 89.63 | 88.52 |
| FUSION (ORIG+PYT) | 90.37 | 85.92 | 92.59 | 90.37 | 74.07 | 93.33 | 87.78 |
| FUSION (ORIG+MAT) | 92.59 | 85.18 | 92.59 | 85.93 | 74.07 | 86.67 | 86.17 |
| FUSION (ORIG+PYT+MAT) | 88.89 | 78.52 | 96.30 | 91.85 | 74.81 | 94.07 | 87.41 |

SVM (PS augmentation, Table III) – PS augmentation, number of augmentations 2 (Subset percentage 50%)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (GEES database) | | | | | |  |
| **151** | **152** | **153** | **154** | **155** | **156** | **AVG** |
| NO AUG (ORIGINAL) | 91.85 | 89.63 | 95.56 | 88.15 | 83.70 | 86.67 | 89.26 |
| AUG (PYTHON) | 88.89 | 80.74 | 97.04 | 89.63 | 85.18 | 92.59 | 89.01 |
| AUG (MATLAB) | 89.63 | 92.59 | 92.59 | 85.18 | 81.48 | 89.63 | 88.52 |
| FUSION (ORIG+PYT) | 89.63 | 81.48 | 96.30 | 88.89 | 84.44 | 97.04 | 89.63 |
| FUSION (ORIG+MAT) | 88.89 | 93.33 | 93.33 | 87.41 | 82.96 | 92.59 | 89.75 |
| FUSION (ORIG+PYT+MAT) | 90.37 | 85.18 | 92.59 | 91.85 | 86.67 | 96.30 | 90.49 |

SVM (PS augmentation, Table III) – PS augmentation, number of augmentations 1 (FULL)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (GEES database) | | | | | |  |
| **151** | **152** | **153** | **154** | **155** | **156** | **AVG** |
| NO AUG (ORIGINAL) | 91.11 | 93.33 | 96.30 | 86.67 | 90.37 | 85.91 | 90.61 |
| AUG (PYTHON) | 88.89 | 82.96 | 96.30 | 85.18 | 93.33 | 92.59 | 89.88 |
| AUG (MATLAB) | 89.63 | 88.89 | 97.04 | 89.63 | 86.67 | 92.59 | 90.74 |
| FUSION (ORIG+PYT) | 90.37 | 90.37 | 96.30 | 87.41 | 89.63 | 92.59 | 91.11 |
| FUSION (ORIG+MAT) | 88.89 | 92.59 | 96.30 | 89.63 | 90.37 | 94.81 | 92.10 |
| FUSION (ORIG+PYT+MAT) | 89.63 | 93.33 | 96.30 | 91.85 | 92.59 | 94.07 | 92.96 |

CNN (PS augmentation, Table IV) – PS augmentation, number of augmentations 10 (Subset percentage 10%)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (GEES database) | | | | | |  |
| **151** | **152** | **153** | **154** | **155** | **156** | **AVG** |
| NO AUG (ORIGINAL) | 70.67 | 38 | 68 | 70.67 | 40.67 | 67.33 | 59.22 |
| AUG (PYTHON) | 74 | 55.33 | 84 | 71.33 | 52 | 70 | 67.78 |
| AUG (MATLAB) | 76 | 52 | 77.33 | 76 | 53.33 | 69.33 | 67.33 |
| FUSION (ORIG+PYT) | 76.67 | 52 | 87.33 | 74.67 | 52.67 | 70 | 68.89 |
| FUSION (ORIG+MAT) | 76.67 | 46.67 | 80 | 78 | 51.33 | 69.33 | 67 |
| FUSION (ORIG+PYT+MAT) | 78 | 54 | 80.67 | 78.67 | 51.33 | 68.67 | 68.56 |

CNN (PS augmentation, Table IV) – PS augmentation, number of augmentations 5 (Subset percentage 20%)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (GEES database) | | | | | |  |
| **151** | **152** | **153** | **154** | **155** | **156** | **AVG** |
| NO AUG (ORIGINAL) | 76 | 36.67 | 73.33 | 78.67 | 48 | 74 | 64.44 |
| AUG (PYTHON) | 82.67 | 56 | 86.67 | 81.33 | 52 | 72.67 | 71.89 |
| AUG (MATLAB) | 83.33 | 58 | 82 | 80.67 | 60.67 | 81.33 | 74.33 |
| FUSION (ORIG+PYT) | 85.33 | 55.33 | 88 | 82 | 53.33 | 75.33 | 73.22 |
| FUSION (ORIG+MAT) | 86 | 56 | 83.33 | 83.33 | 57.33 | 80 | 74.33 |
| FUSION (ORIG+PYT+MAT) | 82.67 | 60.67 | 84 | 84.67 | 55.33 | 78.67 | 74.33 |

CNN (PS augmentation, Table IV) – PS augmentation, number of augmentations 2 (Subset percentage 50%)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (GEES database) | | | | | |  |
| **151** | **152** | **153** | **154** | **155** | **156** | **AVG** |
| NO AUG (ORIGINAL) | 87.33 | 51.33 | 78.67 | 85.33 | 48.67 | 79.33 | 71.78 |
| AUG (PYTHON) | 83.33 | 66.67 | 86.67 | 86 | 54.67 | 80 | 76.22 |
| AUG (MATLAB) | 84 | 62.67 | 84.67 | 84 | 57.33 | 80.67 | 75.56 |
| FUSION (ORIG+PYT) | 88 | 67.33 | 89.33 | 88.67 | 62.67 | 84.67 | 80.11 |
| FUSION (ORIG+MAT) | 85.33 | 60 | 89.33 | 85.33 | 63.33 | 80.67 | 77.33 |
| FUSION (ORIG+PYT+MAT) | 88.67 | 64.67 | 93.33 | 90 | 62 | 83.33 | 80.33 |

CNN (PS augmentation, Table IV) – PS augmentation, number of augmentations 1 (FULL)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| AUGMENTATION | SPEAKER (GEES database) | | | | | |  |
| **151** | **152** | **153** | **154** | **155** | **156** | **AVG** |
| NO AUG (ORIGINAL) | 89.33 | 56.67 | 82.67 | 87.93 | 55.4 | 82.07 | 75.68 |
| AUG (PYTHON) | 88.67 | 64.67 | 87.33 | 85.33 | 61.33 | 78.67 | 77.67 |
| AUG (MATLAB) | 88 | 62.67 | 88 | 85.33 | 66 | 80 | 78.33 |
| FUSION (ORIG+PYT) | 91.33 | 66 | 91.33 | 87.33 | 57.33 | 78.67 | 78.67 |
| FUSION (ORIG+MAT) | 91.33 | 66.67 | 90.67 | 90 | 61.33 | 84 | 80.67 |
| FUSION (ORIG+PYT+MAT) | 92.67 | 66.67 | 93.33 | 89.33 | 61.33 | 86.67 | 81.67 |