**Supplementary Material**

**Tables**

**TABLE S1.** Top 10 highest spearman rank correlations between audio features calculated per speech task and NPI subscales. Results are split up by gender.

(a) (Female; Anxiety)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Positive story** |  |  | **Negative story** |  |  |
| **Feature** | **Correlation** | **p value** | **Feature** | **Correlation** | **p value** |
| sound\_to\_noise\_ratio | 0.3469641 | 0.0007031 | peak\_frequency | -0.4433038 | 9.622e-06 |
| peak\_frequency | -0.3012164 | 0.003526 | amplitude\_mav | -0.4230408 | 2.655e-05 |
| amplitude\_kurtosis | 0.2851256 | 0.005872 | total\_power | -0.4223768 | 2.742e-05 |
| amplitude\_mav | -0.2809288 | 0.006676 | mean\_power | -0.4223768 | 2.742e-05 |
| total\_power | -0.2690185 | 0.009513 | amplitude\_fourth\_moment | -0.3832605 | 0.0001627 |
| mean\_power | -0.2690185 | 0.009513 | max\_amplitude | -0.3171703 | 0.002066 |
| amplitude\_fourth\_moment | -0.2468373 | 0.01769 | sound\_to\_noise\_ratio | 0.2830272 | 0.006262 |
| speech\_ratio | -0.2232307 | 0.03244 | mean\_f0 | -0.2781462 | 0.007262 |
| max\_amplitude | -0.1895974 | 0.07027 | speech\_ratio | -0.273535 | 0.008332 |
| total\_phonation\_time | -0.1268491 | 0.2282 | amplitude\_kurtosis | 0.258846 | 0.01272 |

(b) (Male; Anxiety)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Positive story** |  |  | **Negative story** |  |  |
| **Feature** | **Correlation** | **p value** | **Feature** | **Correlation** | **p value** |
| sound\_to\_noise\_ratio | 0.3396247 | 0.01697 | amplitude\_kurtosis | 0.3433922 | 0.01571 |
| espinola\_zero\_crossing\_metric | -0.1831235 | 0.2079 | sound\_to\_noise\_ratio | 0.3172379 | 0.02635 |
| total\_phonation\_time | -0.1097501 | 0.4528 | amplitude\_mav | -0.2490399 | 0.08442 |
| power\_spectrum\_ratio | 0.1057096 | 0.4697 | speech\_ratio | -0.2163333 | 0.1354 |
| total\_power | -0.09555357 | 0.5137 | number\_of\_pauses | -0.2162034 | 0.1357 |
| mean\_power | -0.09555357 | 0.5137 | mean\_f0 | -0.1889312 | 0.1936 |
| speech\_ratio | 0.09511676 | 0.5156 | peak\_frequency | -0.1708498 | 0.2405 |
| mean\_f0 | -0.07978293 | 0.5858 | total\_power | -0.164516 | 0.2586 |
| local\_absolute\_jitter | 0.06655989 | 0.6495 | mean\_power | -0.164516 | 0.2586 |
| average\_amplitude\_change | -0.06585006 | 0.653 | total\_phonation\_time | -0.159711 | 0.273 |

(c) (Female; Depression)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Positive story** |  |  | **Negative story** |  |  |
| **Feature** | **Correlation** | **p value** | **Feature** | **Correlation** | **p value** |
| peak\_frequency | -0.2291068 | 0.02804 | max\_amplitude | -0.292862 | 0.004611 |
| amplitude\_fourth\_moment | -0.2163193 | 0.03835 | peak\_frequency | -0.2623844 | 0.01151 |
| total\_power | -0.2116438 | 0.04284 | amplitude\_fourth\_moment | -0.260786 | 0.01205 |
| mean\_power | -0.2116438 | 0.04284 | total\_power | -0.2544122 | 0.01439 |
| amplitude\_mav | -0.1882766 | 0.07229 | mean\_power | -0.2544122 | 0.01439 |
| max\_amplitude | -0.1748197 | 0.09556 | amplitude\_mav | -0.2281278 | 0.02873 |
| amplitude\_third\_moment | -0.1557183 | 0.1383 | local\_jitter | 0.1580161 | 0.1325 |
| sound\_to\_noise\_ratio | 0.1425711 | 0.1752 | ppq5\_jitter | 0.1381954 | 0.1889 |
| mean\_f0 | -0.1225804 | 0.2444 | rap\_jitter | 0.1316518 | 0.211 |
| espinola\_zero\_crossing\_metric | -0.0778772 | 0.4606 | ddp\_jitter | 0.1316518 | 0.211 |

(d) (Male; Depression)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Positive story** |  |  | **Negative story** |  |  |
| **Feature** | **Correlation** | **p value** | **Feature** | **Correlation** | **p value** |
| average\_amplitude\_change | 0.2030018 | 0.1618 | average\_amplitude\_change | 0.2699959 | 0.06063 |
| harmonics\_to\_noise\_ratio | 0.1820364 | 0.2106 | amplitude\_fourth\_moment | 0.1450946 | 0.3199 |
| local\_jitter | -0.1503211 | 0.3026 | peak\_frequency | 0.1181306 | 0.4189 |
| ppq5\_jitter | -0.1214566 | 0.4058 | local\_absolute\_jitter | -0.1035796 | 0.4788 |
| local\_absolute\_jitter | -0.1152798 | 0.4303 | mean\_f0 | 0.09952806 | 0.4962 |
| rap\_jitter | -0.1086873 | 0.4573 | power\_spectrum\_ratio | 0.0969277 | 0.5076 |
| ddp\_jitter | -0.1086873 | 0.4573 | local\_jitter | 0.08112944 | 0.5795 |
| amplitude\_fourth\_moment | 0.1041735 | 0.4763 | ppq5\_jitter | 0.08053552 | 0.5823 |
| total\_phonation\_time | -0.09591804 | 0.5121 | total\_power | 0.07394301 | 0.6136 |
| max\_amplitude | 0.09455202 | 0.5181 | mean\_power | 0.07394301 | 0.6136 |

(e) (Female; Apathy)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Positive story** |  |  | **Negative story** |  |  |
| **Feature** | **Correlation** | **p value** | **Feature** | **Correlation** | **p value** |
| speech\_ratio | -0.4649382 | 3.024e-06 | speech\_ratio | -0.4422672 | 1.015e-05 |
| amplitude\_kurtosis | 0.4122244 | 4.447e-05 | number\_of\_pauses | -0.3738093 | 0.0002421 |
| number\_of\_pauses | -0.4005278 | 7.617e-05 | amplitude\_mav | -0.3416138 | 0.0008601 |
| total\_phonation\_time | -0.3871873 | 0.0001374 | total\_phonation\_time | -0.3212951 | 0.001791 |
| amplitude\_mav | -0.2895547 | 0.005117 | total\_power | -0.3045195 | 0.003164 |
| peak\_frequency | -0.2145304 | 0.04002 | mean\_power | -0.3045195 | 0.003164 |
| total\_power | -0.199052 | 0.05714 | harmonics\_to\_noise\_ratio | -0.3018382 | 0.003455 |
| mean\_power | -0.199052 | 0.05714 | peak\_frequency | -0.2815369 | 0.006554 |
| harmonics\_to\_noise\_ratio | -0.1656924 | 0.1145 | amplitude\_kurtosis | 0.2753473 | 0.007896 |
| amplitude\_fourth\_moment | -0.123723 | 0.24 | amplitude\_fourth\_moment | -0.250406 | 0.01606 |

(f) (Male;Apathy)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Positive story** |  |  | **Negative story** |  |  |
| **Feature** | **Correlation** | **p value** | **Feature** | **Correlation** | **p value** |
| speech\_ratio | -0.5011488 | 0.0002446 | speech\_ratio | -0.5768886 | 1.433e-05 |
| number\_of\_pauses | -0.3865379 | 0.00608 | amplitude\_kurtosis | 0.455276 | 0.001013 |
| amplitude\_kurtosis | 0.3169775 | 0.02648 | number\_of\_pauses | -0.4269726 | 0.002216 |
| total\_phonation\_time | -0.3050516 | 0.03306 | total\_phonation\_time | -0.3781761 | 0.00738 |
| amplitude\_mav | -0.2994025 | 0.03662 | amplitude\_mav | -0.2946426 | 0.03987 |
| peak\_frequency | -0.2489267 | 0.08457 | peak\_frequency | -0.2666586 | 0.06401 |
| total\_power | -0.2485083 | 0.08511 | total\_power | -0.2493975 | 0.08396 |
| mean\_power | -0.2485083 | 0.08511 | mean\_power | -0.2493975 | 0.08396 |
| espinola\_zero\_crossing\_metric | -0.1929972 | 0.184 | local\_jitter | 0.1978233 | 0.173 |
| mean\_f0 | 0.1854857 | 0.202 | mean\_f0 | 0.181538 | 0.2119 |

**TABLE S2.** Top 10 highest spearman rank partial correlations between audio features calculated per speech task and NPI subscales, corrected for MMSE. Results are split up by gender.

(a) (Female; Anxiety)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Positive story** |  |  | **Negative story** |  |  |
| **Feature** | **Correlation** | **p value** | **Feature** | **Correlation** | **p value** |
| sound\_to\_noise\_ratio | 0.33073 | 0.00136 | mean\_f0 | -0.36523 | 0.00037 |
| peak\_frequency | -0.20172 | 0.05518 | peak\_frequency | -0.35242 | 0.00061 |
| amplitude\_fourth\_moment | -0.17364 | 0.09975 | total\_power | -0.32664 | 0.00158 |
| total\_power | -0.17349 | 0.10004 | mean\_power | -0.32664 | 0.00158 |
| mean\_power | -0.17349 | 0.10004 | amplitude\_mav | -0.32015 | 0.00198 |
| amplitude\_kurtosis | 0.16728 | 0.11298 | amplitude\_fourth\_moment | -0.29205 | 0.00497 |
| amplitude\_mav | -0.16361 | 0.12124 | local\_absolute\_jitter | 0.28813 | 0.00561 |
| mean\_f0 | -0.12298 | 0.24548 | sound\_to\_noise\_ratio | 0.27374 | 0.00865 |
| local\_absolute\_jitter | 0.11589 | 0.27400 | max\_amplitude | -0.23753 | 0.02339 |
| max\_amplitude | -0.11202 | 0.29043 | ppq5\_jitter | 0.23121 | 0.02744 |

(b) (Male; Anxiety)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Positive story** |  |  | **Negative story** |  |  |
| **Feature** | **Correlation** | **p value** | **Feature** | **Correlation** | **p value** |
| sound\_to\_noise\_ratio | 0.39219 | 0.00583 | sound\_to\_noise\_ratio | 0.40730 | 0.00406 |
| espinola\_zero\_crossing\_metric | -0.17880 | 0.22400 | amplitude\_kurtosis | 0.27846 | 0.05531 |
| power\_spectrum\_ratio | 0.13428 | 0.36285 | mean\_f0 | -0.23137 | 0.11358 |
| speech\_ratio | 0.13293 | 0.36776 | amplitude\_mav | -0.22492 | 0.12430 |
| ppq5\_jitter | -0.11695 | 0.42858 | ppq5\_jitter | -0.17848 | 0.22484 |
| total\_power | -0.10299 | 0.48604 | peak\_frequency | -0.16178 | 0.27196 |
| mean\_power | -0.10299 | 0.48604 | number\_of\_pauses | -0.15929 | 0.27950 |
| mean\_f0 | -0.10192 | 0.49062 | total\_power | -0.15637 | 0.28852 |
| average\_amplitude\_change | -0.08493 | 0.56600 | mean\_power | -0.15637 | 0.28852 |
| amplitude\_fourth\_moment | -0.07194 | 0.62704 | speech\_ratio | -0.14818 | 0.31482 |

(c) (Female; Depression)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Positive story** |  |  | **Negative story** |  |  |
| **Feature** | **Correlation** | **p value** | **Feature** | **Correlation** | **p value** |
| peak\_frequency | -0.18970 | 0.07170 | max\_amplitude | -0.26266 | 0.01189 |
| amplitude\_fourth\_moment | -0.18571 | 0.07800 | amplitude\_fourth\_moment | -0.22337 | 0.03331 |
| total\_power | -0.17313 | 0.10076 | peak\_frequency | -0.22293 | 0.03367 |
| mean\_power | -0.17313 | 0.10076 | total\_power | -0.21372 | 0.04193 |
| max\_amplitude | -0.14265 | 0.17737 | mean\_power | -0.21372 | 0.04193 |
| amplitude\_mav | -0.14058 | 0.18381 | amplitude\_mav | -0.18209 | 0.08408 |
| amplitude\_third\_moment | -0.13400 | 0.20540 | amplitude\_third\_moment | -0.17726 | 0.09277 |
| mean\_f0 | -0.12836 | 0.22530 | local\_jitter | 0.15316 | 0.14722 |
| sound\_to\_noise\_ratio | 0.12835 | 0.22533 | local\_absolute\_jitter | 0.14442 | 0.17200 |
| average\_amplitude\_change | -0.09983 | 0.34644 | ppq5\_jitter | 0.13458 | 0.20342 |

(d) (Male; Depression)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Positive story** |  |  | **Negative story** |  |  |
| **Feature** | **Correlation** | **p value** | **Feature** | **Correlation** | **p value** |
| harmonics\_to\_noise\_ratio | 0.22161 | 0.13008 | average\_amplitude\_change | 0.28733 | 0.04768 |
| average\_amplitude\_change | 0.19633 | 0.18108 | amplitude\_fourth\_moment | 0.14453 | 0.32705 |
| local\_jitter | -0.17628 | 0.23070 | local\_absolute\_jitter | -0.13237 | 0.36979 |
| ppq5\_jitter | -0.15358 | 0.29733 | peak\_frequency | 0.12938 | 0.38080 |
| rap\_jitter | -0.13196 | 0.37127 | speech\_ratio | 0.11710 | 0.42801 |
| ddp\_jitter | -0.13196 | 0.37127 | power\_spectrum\_ratio | 0.10835 | 0.46352 |
| local\_absolute\_jitter | -0.12744 | 0.38802 | number\_of\_pauses | 0.10633 | 0.47196 |
| amplitude\_fourth\_moment | 0.09941 | 0.50142 | amplitude\_kurtosis | -0.09845 | 0.50561 |
| power\_spectrum\_ratio | 0.09347 | 0.52743 | sound\_to\_noise\_ratio | 0.09809 | 0.50713 |
| max\_amplitude | 0.09334 | 0.52804 | total\_power | 0.08377 | 0.57134 |

(e) (Female; Apathy)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Positive story** |  |  | **Negative story** |  |  |
| **Feature** | **Correlation** | **p value** | **Feature** | **Correlation** | **p value** |
| amplitude\_kurtosis | 0.22271 | 0.03385 | harmonics\_to\_noise\_ratio | -0.24608 | 0.01871 |
| speech\_ratio | -0.19707 | 0.06116 | speech\_ratio | -0.22413 | 0.03270 |
| total\_phonation\_time | -0.17709 | 0.09311 | ppq5\_jitter | 0.18561 | 0.07816 |
| number\_of\_pauses | -0.16877 | 0.10978 | rap\_jitter | 0.17773 | 0.09191 |
| sound\_to\_noise\_ratio | -0.16037 | 0.12889 | ddp\_jitter | 0.17773 | 0.09191 |
| harmonics\_to\_noise\_ratio | -0.12995 | 0.21955 | local\_jitter | 0.16416 | 0.11998 |
| mean\_f0 | 0.10509 | 0.32151 | sound\_to\_noise\_ratio | -0.13812 | 0.19168 |
| amplitude\_mav | -0.05878 | 0.57995 | number\_of\_pauses | -0.13435 | 0.20419 |
| ppq5\_jitter | 0.05321 | 0.61643 | espinola\_zero\_crossing\_metric | -0.11075 | 0.29597 |
| power\_spectrum\_ratio | -0.05298 | 0.61795 | total\_phonation\_time | -0.10475 | 0.32307 |

(f) (Apathy, Male)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Positive story** |  |  | **Negative story** |  |  |
| **Feature** | **Correlation** | **p value** | **Feature** | **Correlation** | **p value** |
| speech\_ratio | -0.51070 | 0.00021 | speech\_ratio | -0.51766 | 0.00016 |
| amplitude\_mav | -0.33569 | 0.01967 | number\_of\_pauses | -0.35981 | 0.01201 |
| number\_of\_pauses | -0.30012 | 0.03821 | total\_phonation\_time | -0.35587 | 0.01305 |
| total\_power | -0.29197 | 0.04405 | amplitude\_kurtosis | 0.35308 | 0.01383 |
| mean\_power | -0.29197 | 0.04405 | peak\_frequency | -0.27395 | 0.05954 |
| peak\_frequency | -0.27980 | 0.05409 | amplitude\_mav | -0.26857 | 0.06493 |
| amplitude\_kurtosis | 0.26409 | 0.06971 | total\_power | -0.25654 | 0.07840 |
| total\_phonation\_time | -0.22924 | 0.11704 | mean\_power | -0.25654 | 0.07840 |
| amplitude\_fourth\_moment | -0.20050 | 0.17180 | amplitude\_fourth\_moment | -0.19747 | 0.17852 |
| average\_amplitude\_change | -0.19865 | 0.17588 | amplitude\_third\_moment | -0.17273 | 0.24038 |

**TABLE S3.** Top 5 highest feature weights selected by SVM regression models in predicting each NPI subscale, seperated for females and males.

1. Anxiety

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Female** | | | |  | **Male** | | | |
| **Feature** | **Task** | **Mean** | **Standard Deviation** |  | **Feature** | **Task** | **Mean** | **Standard Deviation** |
| MFCC 18 | Neg | 0.0192 | 0.00076 |  | MFCC 9 | Neg | 0.0170 | 0.00050 |
| MFCC 4 | Neg | 0.0188 | 0.00061 |  | MFCC 4 | Pos | 0.0154 | 0.00060 |
| Sound-to-noise ratio | Pos | 0.0182 | 0.00070 |  | MFCC 10 | Pos | 0.0146 | 0.00051 |
| Deltas 7 | Pos | 0.0178 | 0.00083 |  | MFCC 16 | Neg | 0.0143 | 0.00075 |
| Delta Deltas 17 | Neg | 0.0172 | 0.00086 |  | MFCC 6 | Neg | 0.0141 | 0.00067 |

1. Depression

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Female** | | | |  | **Male** | | | |
| **Feature** | **Task** | **Mean** | **Standard Deviation** |  | **Feature** | **Task** | **Mean** | **Standard Deviation** |
| Delta Deltas 11 | Neg | 0.0014 | 5.13151e-05 |  | Delta Deltas 5 | Neg | 0.0009 | 6.46563e-05 |
| Deltas 8 | Pos | 0.0013 | 6.30781e-05 |  | Delta Deltas 20 | Neg | 0.0009 | 9.55432e-05 |
| Deltas 9 | Pos | 0.0011 | 5.65088e-05 |  | Amplitude kurtosis | Neg | 0.0008 | 3.10521e-05 |
| Deltas 16 | Neg | 0.0009 | 5.16828e-05 |  | Deltas 1 | Neg | 0.0007 | 7.12870e-05 |
| Deltas 17 | Neg | 0.0009 | 6.49522e-05 |  | Delta deltas 14 | Pos | 0.0007 | 7.22985e-05 |

(c) Apathy

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Female** | | | |  | **Male** | | | |
| **Feature** | **Task** | **Mean** | **Standard Deviation** |  | **Feature** | **Task** | **Mean** | **Standard Deviation** |
| MMSE |  | 0.0035 | 0.00011 |  | Speech ratio | Neg | 0.2203 | 0.09936 |
| Speech ratio | Pos | 0.0021 | 0.00011 |  | MMSE |  | 0.1910 | 0.09611 |
| Speech ratio | Neg | 0.0019 | 0.00012 |  | Speech ratio | Pos | 0.1774 | 0.04123 |
| Harmonics-to-noise ratio | Neg | 0.0018 | 9.59043e-05 |  | Deltas 5 | Neg | 0.1289 | 0.09128 |
| Deltas 2 | Pos | 0.0017 | 8.55444e-05 |  | Number of pauses | Neg | 0.1187 | 0.02705 |