

Supplemental Material S1. In the article, we reported only one of the six components of the NASA-TLX (mental demand) because it seemed the most relevant. Here, we report on the other five components: physical demand, temporal demand, subjective performance, effort, and frustration.

Experiment 1

Physical demand: There was no main effect of pitch conditions ($F(3, 180) = 1.191, p = .314, \omega^2 < 0.001$), no main effect of group ($F(1, 60) = 0.429, p = .515, \omega^2 < 0.001$), and no interaction ($F(3, 180) = 2.417, p = .068, \omega^2 < 0.001$).

Temporal demand: There was no main effect of pitch conditions ($F(3, 180) = 1.583, p = .195, \omega^2 < 0.001$) and no interaction ($F(3, 180) = 0.040, p = .989, \omega^2 < 0.001$), though there was a main effect of group ($F(1, 60) = 5.390, p = .024, \omega^2 = 0.035$). French-speaking participants reported higher temporal demand (mean rating = 5.151, $SD = 1.619$) than English-speaking participants (mean rating = 4.230, $SD = 1.503$).

Subjective performance: There was no main effect of pitch conditions ($F(3, 180) = 0.248, p = .862, \omega^2 < 0.001$), no main effect of group ($F(1, 60) = 2.321, p = .133, \omega^2 = 0.011$), and no interaction ($F(3, 180) = 1.274, p = .285, \omega^2 < 0.001$).

Effort: There was no main effect of pitch conditions ($F(3, 180) = 1.041, p = .376, \omega^2 < 0.001$), no main effect of group ($F(1, 60) = 2.929, p = .092, \omega^2 = 0.016$), and no interaction ($F(3, 180) = 1.115, p = .345, \omega^2 < 0.001$).

Frustration: There was no main effect of pitch conditions ($F(3, 180) = 0.283, p = .838, \omega^2 < 0.001$), no main effect of group ($F(1, 60) = 1.895 \times 10^{-4}, p = .989, \omega^2 < 0.001$), and no interaction ($F(3, 180) = 0.626, p = .599, \omega^2 < 0.001$).

To summarize, there were no interactions with group in any of these ratings, and only one main effect of group seen with temporal demand, in line with that reported in the article for mental demand.

Experiment 2

Physical demand: There was no main effect of timing conditions ($F(3, 186) = 0.615, p = .606, \omega^2 < 0.001$), no main effect of group ($F(1, 62) = 1.447, p = .234, \omega^2 = 0.004$), and no interaction ($F(3, 186) = 0.626, p = .599, \omega^2 < 0.001$).

Temporal demand: There was no main effect of group ($F(1, 62) = 3.728, p = .058, \omega^2 = 0.021$) but a main effect of timing conditions ($F(3, 186) = 49.464, p < .001, \omega^2 = 0.078$) with participants reporting the highest temporal demand in fixed-fast (mean rating = 4.986, $SD = 1.317$), followed by rhythmic (mean rating = 4.534, $SD = 1.296$), jitter (mean rating = 4.306, $SD = 1.286$) and fixed-slow (mean rating = 3.949, $SD = 1.389$). More interestingly, and consistent with Experiment 1, there was also an interaction between timing and group ($F(3, 186) = 5.190, p = .002, \omega^2 = 0.007$).

because English-speaking participants indicated lower temporal demand in the fixed-slow condition than French-speaking participants.

Subjective performance: There was no main effect of group ($F(1, 62) = 0.544, p = .463, \omega^2 < 0.001$) and no interaction ($F(3, 186) = 1.039, p = .376, \omega^2 < 0.001$). The main effect of timing conditions ($F(3, 186) = 25.517, p < .001, \omega^2 = 0.037$) reflected the highest subjective performance in fixed-slow (mean rating = 4.021, $SD = 1.365$), followed by jitter (mean rating = 3.759, $SD = 1.301$), rhythmic (mean rating = 3.742, $SD = 1.267$) and fixed-fast (mean rating = 3.317, $SD = 1.086$).

Effort: There was no main effect of timing conditions ($F(3, 186) = 1.021, p = .385, \omega^2 < 0.001$) and no main effect of group ($F(1, 62) = 0.099, p = .754, \omega^2 < 0.001$), though there was an interaction between timing and group ($F(3, 186) = 2.803, p = .041, \omega^2 = 0.002$) because English-speaking participants indicated lower effort in the fixed-slow condition than French-speaking participants.

Frustration: There was no main effect of group ($F(1, 62) = 0.012, p = .914, \omega^2 < 0.001$) and no interaction ($F(3, 186) = 0.333, p = .801, \omega^2 < 0.001$). The main effect of timing conditions ($F(3, 186) = 17.341, p < .001, \omega^2 = 0.010$) reflected a higher frustration in fixed-fast (mean rating = 3.591, $SD = 1.663$), followed by rhythmic (mean rating = 3.276, $SD = 1.634$), jitter (mean rating = 3.195, $SD = 1.601$) and fixed-slow (mean rating = 3.183, $SD = 1.613$).

To summarize, there were no main effects of group in any of these ratings, and only two interactions with group in temporal demand and effort, suggesting that French-speaking participants found the pace of the task more challenging than English-speaking participants did in the easiest condition (fixed-slow). This result was in line with French-speaking participants being particularly diligent in these tasks, particularly when presentation time allowed it. As for the differences among the four experimental conditions, they were as expected, with fixed-fast being the hardest condition, fixed-slow the easiest, and jitter and rhythmic at intermediate levels of difficulty.

Experiment 3

Physical demand: There was no main effect of pitch ($F(1, 64) = 2.579, p = .113, \omega^2 < 0.001$), no main effect of timing ($F(1, 64) = 0.520, p = .474, \omega^2 < 0.001$), no main effect of group ($F(1, 64) = 1.227, p = .272, \omega^2 = 0.002$), no interaction between pitch and group ($F(1, 64) = 0.259, p = .613, \omega^2 < 0.001$), no interaction between timing and group ($F(1, 64) = 0.025, p = .876, \omega^2 < 0.001$), no interaction between pitch and timing ($F(1, 64) = 0.473, p = .494, \omega^2 < 0.001$), and no three-way interaction ($F(1, 64) = 0.065, p = .800, \omega^2 < 0.001$).

Temporal demand: There was no main effect of group ($F(1, 64) = 0.736, p = .394, \omega^2 = 0.002$), no interaction between pitch and group ($F(1, 64) = 1.433, p = .236, \omega^2 < 0.001$), no interaction between timing and group ($F(1, 64) = 0.527, p = .471, \omega^2 < 0.001$), no interaction between pitch and timing ($F(1, 64) = 0.096, p = .757, \omega^2 < 0.001$), and no three-way interaction ($F(1, 64) = 0.977, p = .327, \omega^2 < 0.001$). However, there was a main effect of pitch ($F(1, 64) = 6.218, p = .015, \omega^2 = 0.002$) with participants reporting higher temporal demand in the flat conditions (mean rating = 4.265, $SD = 1.601$) compared to arpeggio conditions (mean rating = 4.141, $SD = 1.565$), as well as a main effect of timing ($F(1, 64) = 6.226, p = .014, \omega^2 = 0.001$) with participants

reporting higher temporal demand in the fixed conditions (mean rating = 4.274, $SD = 1.612$) compared to spaced conditions (mean rating = 4.132, $SD = 1.554$).

Subjective performance: There was no main effect of timing ($F(1, 64) = 3.419$, $p = .069$, $\omega^2 < 0.001$), no main effect of group ($F(1, 64) = 3.270$, $p = .075$, $\omega^2 = 0.017$), no interaction between pitch and group ($F(1, 64) = 5.786 \times 10^{-4}$, $p = .965$, $\omega^2 < 0.001$), no interaction between timing and group ($F(1, 64) = 0.733$, $p = .395$, $\omega^2 < 0.001$), no interaction between pitch and timing ($F(1, 64) = 1.288$, $p = .261$, $\omega^2 < 0.001$), and no three-way interaction ($F(1, 64) = 0.621$, $p = .434$, $\omega^2 < 0.001$). However, there was a main effect of pitch ($F(1, 64) = 10.046$, $p = .002$, $\omega^2 = 0.008$) with participants reporting higher subjective performance in arpeggio conditions (mean rating = 3.981, $SD = 1.191$) compared to flat conditions (mean rating = 3.854, $SD = 1.202$).

Effort: There was no main effect of pitch ($F(1, 64) = 3.341$, $p = .072$, $\omega^2 < 0.001$), no main effect of group ($F(1, 64) = 2.493$, $p = .119$, $\omega^2 = 0.011$), no interaction between pitch and group ($F(1, 64) = 0.251$, $p = .618$, $\omega^2 < 0.001$), no interaction between pitch and timing ($F(1, 64) = 3.375$, $p = .071$, $\omega^2 < 0.001$), and no three-way interaction ($F(1, 64) = 0.757$, $p = .388$, $\omega^2 < 0.001$). However, there was a main effect of timing ($F(1, 64) = 5.234$, $p = .025$, $\omega^2 < 0.001$) with participants reporting higher effort for fixed conditions (mean rating = 5.204, $SD = 1.228$) compared to spaced conditions (mean rating = 5.130, $SD = 1.248$). There was also an interaction between timing and group ($F(1, 64) = 7.337$, $p = .009$, $\omega^2 = 0.001$) because English-speaking participants reported lower effort on spaced conditions than French-speaking participants.

Frustration: There was no main effect of timing ($F(1, 64) = 0.916$, $p = .342$, $\omega^2 < 0.001$), no main effect of group ($F(1, 64) = 1.949$, $p = .168$, $\omega^2 = 0.007$), no interaction between pitch and group ($F(1, 64) = 0.015$, $p = .904$, $\omega^2 < 0.001$), no interaction between timing and group ($F(1, 64) = 0.626$, $p = .904$, $\omega^2 < 0.001$), no interaction between pitch and timing ($F(1, 64) = 1.036$, $p = .313$, $\omega^2 < 0.001$), and no three-way interaction ($F(1, 64) = 8.059 \times 10^{-6}$, $p = .998$, $\omega^2 < 0.001$). However, there was a main effect of pitch condition ($F(1, 64) = 4.954$, $p = .030$, $\omega^2 = 0.002$), with participants reporting lower frustration in the arpeggio conditions (mean rating = 3.200, $SD = 1.655$) than in the flat conditions (mean rating = 3.259, $SD = 1.677$).

To summarize, there were some differences expected of grouping that indicated it was less effortful when lists were grouped rather than flat/fixed. There were no main effects of group in any of these ratings, and only one interaction with group seen with effort (group x timing), suggesting that French-speaking participants spent more effort than English-speaking participants when grouping could be applied in time. These traits are therefore globally in line with what is reported in the main article.