Figure captions

**Figure 1 (left panel):** Averaged tap-tone asynchronies for tone sequences containing negative shifts (tone presented 90 ms earlier than expected: filled circle) and positive shifts (tone presented 90 ms later than expected: unfilled circle). On the x-axis, $T_0$ denotes the stimulus where the shift occurred. Four regular tone sequences before ($T_4$ to $T_1$) and after ($T_1$ to $T+4$) the shift are shown. Figure 1 (right panel): The identical data were transformed to show normalized asynchronies following a shift ($T_0$), to compare the error correction performance between negative and positive shift conditions. On the y-axis, '0' indicates the baseline negative mean asynchrony (average of $T_4$ to $T_1$), and '1' on the y-axis shows the maximum deviance from the baseline owing to the shift. Positive shifts were corrected faster with a degree of over-correction (unfilled circle), compared with negative shifts (filled circle), [$p < .05$]. Error bars represent standard error of mean.

**Figure 2:** Grand averaged stimulus-locked ERPs to all 4 conditions from FCz for illustration purpose only. ERPs were time-locked to $T_2$ (at 0 ms). These macro-epochs contain preceding tones ($T_2$ & $T_1$), a tone subject to a ±90 ms time-shift ($T_0$), and 4 subsequent tones ($T+1$ to $T+4$). Condition labels indicate the shift direction of $T_0$ (-ve shift: 90 ms earlier than expected or +ve shift: 90 ms later than expected).

**Figure 3 (upper panel):** Grand averaged ERPs from FCz showing stimulus-locked epochs to the shift position $T_0$ (at 0 ms on the x-axis) for listening and tapping conditions of both shift directions. A significant 2-way interaction between Condition (listening vs. tapping) and ShiftDirection (negative vs. positive) was identified in 2 time windows (shaded boxes: N1 around 100 ms [$F(1,14) = 19.77$, $p < .001$] and N2 around 300 ms [$F(1,14) = 15.06$, $p < .001$]). Figure 3 (lower panel): Topographic maps for each
condition for each time window, and their corresponding significance maps were shown (at 119 ms and 316 ms). Note that warmer colors represent positivity.

**Figure 4:** Grand averaged ERPs from FCz showing stimulus-locked epochs, time-locked to *T-1* or *T0* (0 ms on the x-axis) for tapping negative and positive conditions. ERPs were relative to the baseline period from -50 to 0 ms. A significant 2-way interaction Position (*T-1* vs. *T0*) and ShiftDirection (negative vs. positive) was identified in 2 time windows (shaded boxes: N1 around 100 ms [$F(1,14) = 31.55, p < .001$] and N2 around 300 ms [$F(1,14) = 25.13, p < .001$]).

**Figure 5 (upper panel):** Grand averaged ERPs from FCz, showing response-locked epochs, time-locked to the tap-onset for *T-1* or *T0* stimulus (at 0 ms on the x-axis) for tapping negative and positive conditions. ERPs were relative to the baseline period from -50 to 0 ms. No significant 2-way interaction was identified between Position (*T-1* vs. *T0*) and ShiftDirection (negative vs. positive). The shaded box (356-408 ms) indicates the significant window of ShiftDirection main effect (tapping negative condition > tapping positive condition). It was most significant at 374 ms [$F(1,14) = 6.22, p < .05$]. **Figure 5 (lower panel):** Topographic maps showing activity at 374 ms. It compares tapping negative and tapping positive conditions at *T0* only. Note that warmer colors represent positivity.

**Figure 6 (left panel):** The peak amplitude and latency of CNV-like negativity for each participant. In the tapping negative condition, there was a significant positive correlation between the CNV-like negativity peak latency and the normalized error correction performance at *T+1* (i.e., the earlier the peak, the better the error correction performance with the negative shifts) [$\tau(15) = .569, p = .027$]. **Figure 6 (right panel):** In the tapping positive condition, there was a trend level of negative correlation between
the CNV-like negativity peak latency and the normalized error correction performance at 
$T+1$ (i.e., the later the peak, the better the error correction performance with the positive 
shifts) [$r(15) = -.439, p = .10]$. 
Figure 1
Figure 2
Figure 3

[Graph showing EEG potential over time with different conditions labeled: listening -ve T0, listening +ve T0, tapping -ve T0, tapping +ve T0. Two heat maps display tapping responses with p-values and significance marked by red dots.]
Figure 5
Figure 6

![Graph showing the relationship between CNV-like negativity latency and normalized error correction performance at T+1. The left graph shows a positive correlation, while the right graph shows a negative correlation.](image)