Do bimodally presented sequences in the sequence production task offer a benefit in performance and task-specific learning as compared to unimodally presented stimuli?

Bart E.G. Hilhorst, Rob H. J. van der Lubbe, and E. De Kleine

University of Twente

Author Details

Bart Hilhorst, Department of Cognitive Psychology and Ergonomics, University of Twente.

Correspondence concerning this article should be addressed to Rob H. J. van der Lubbe, Department of Cognitive Psychology and Ergonomics, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands. E-mail: r.h.j.vanderlubbe@utwente.nl

Abstract

This study investigated if a bimodal format of stimulus presentation in a sequence production task resulted in improved performance as compared to unimodal stimulus presentation, due to the invocation of visual and auditory working memory, and/or the presence of a sensory integration mechanism resulting in an improved consolidation of sequence information in working memory. The effect of practice on task-specific learning was investigated as well. Participants were presented sequences of four stimuli using unimodal (visual or auditory) or bimodal (audiovisual) stimuli. After a brief preparation period, a Go/Nogo signal was given signaling the participant to either respond by pressing the corresponding keys on the keyboard, or to withhold a response. Behavioral measures as well as the contingent negative variation (CNV) were analyzed. The results showed no beneficial effects for bimodal as compared to visual sequences, while the auditory condition performed worse. We found that the auditory CNV was less pronounced overall, corresponding with the poorer performance in this condition. Furthermore, a session effect was found at C1, suggesting improvement of working memory use and increased general motor preparation as a result of practice. Furthermore, we found that practice resulted in a decrease of the general CNV for bimodal stimuli, whereas the visual and auditory CNV increased. This could indicate that practice leads to a different encoding for bimodal sequences, albeit without beneficial effects on performance.

Keywords: working memory, sensory integration, CNV, DSP-task, modality

One of the things that one has to learn when playing the piano, is the ability to process novel sequences of notes and play these sequences correctly. Typically this ability is developed by performing a variety of musical exercises, such as playing a melody from a sheet of paper or by replaying a melody by ear. Immediate auditory feedback is provided by the piano when pressing a key and in a seemingly automatic fashion musicians become more adept at finding the right keys when playing musical pieces. Over time, musicians become skilled at replaying melodies by sight-reading or by ear alone. Although these activities are different from a perceptual point of view, they both lead to the same result: a correctly played musical piece. In today's modern world, a lot of emphasis is placed on efficiency when it comes to performance and learning of new tasks. Hence one can wonder if a particular sensory modality is most suited for correctly learning and performing such sequence-related tasks. And perhaps more interestingly, what about using both senses simultaneously?

In a task such as playing the piano, sequences of information are stored in working memory after receiving stimulus input from the senses such as reading notes from a sheet of paper. The sequence information in working memory is then used to program the appropriate motor sequence that results in pressing the appropriate keys on the piano. Besides a unimodal format of presenting the sequence information, it is also possible to present sequence information bimodally. Using audiovisual stimuli rather than either visual or auditory stimuli could provide a beneficial effect that would result in improved performance and improved task-specific learning. On the basis of current literature, two reasons may be given why a bimodal presentation of sequence information could be beneficial as compared to a unimodal format of presentation. First, different types of working memory may be involved when storing information from different modalities (Baddeley & Hitch, 1974), meaning that a bimodal format of stimulus presentation employs multiple working memory components. Second, a mechanism known as sensory integration may be involved during stimulus presentation, which could result in a better consolidation of sequence information in working memory. In this study, we investigated if a bimodal presentation of sequence information provided a performance advantage as compared to unimodally presented sequence information. Furthermore, we were interested if a bimodal sequence presentation would lead to a better performance as a result of practice. Participants performed a sequence production task featuring unimodal and bimodal sequences, during which behavioral data as well as cortical activity were recorded. Behavioral data was used to gauge task performance and task specific learning, while the EEG data was analyzed to locate the cortical regions associated with the potential benefits from bimodal sequence presentation. In the following section we will discuss working memory and sensory integration in more detail.

Our working memory is capable of holding a limited amount of information active for a short duration. In a study by Baddeley and Hitch (1974), it was found that performance in two simultaneous tasks was nearly as efficient as performance of the tasks when performed separately, but only when these two tasks required the use of different perceptual domains (e.g. visual vs. verbal information). Performance suffered when the two simultaneous tasks were both performed in the same perceptual domain. Two different working memory systems have been distinguished for sensory input: a visual/visual-spatial and an auditory/verbal component. Typically the modality of the stimulus determines which working memory component is used. For example, the visual-spatial working memory is employed in the Corsi-block spatial span task due to the spatial configuration of the visual stimuli (Fischer, 2001). There is evidence that bimodally presented information can lead to improved performance in working memory tasks due to a dual representation in working memory. A study by Goolkasian and Foos (2005) examined memory performance in a task where a number of items had to be held in memory while a distractor math task was performed. Semantic items were presented by picture, spoken word, printed word or a combination of two formats. It was found that memory performance was better when items had been presented bimodally rather than unimodally, suggesting that bimodally presented information can lead to improved working memory performance. Another study used non-semantic stimuli and found a beneficial effect of bimodal stimuli on memory performance as well (Santangelo, Mastroberardino, Botta, Marucci, & Belardinelli, 2006). In their study, a stream of non-semantic unimodal (visual Chinese ideograms or auditory novel musical fragments) or bimodal (audiovisual) items was presented. Participants had to indicate if a presented item was seen two positions earlier in the item stream. A significant improvement of performance for bimodal stimuli was found as compared to unimodal stimuli. As sequence of stimuli can be considered as a stream of non-semantic information, it may be possible that this principle is also applicable to the storage of bimodally presented sequence information.

Sensory integration is the process in which sensory information in different modalities is unified in order to represent a singular event. Research shows that sensory integration of multimodal stimuli poses behavioral advantages when compared to unimodal stimuli (Molholm et al., 2002; Sella, Reiner, & Pratt, 2013; Talsma, Doty, & Woldorff, 2007). Molholm et al. (2002) found that simultaneous presentation of bimodal (audiovisual) stimuli in a simple reaction-time task resulted in faster reaction times than when stimuli were presented unimodally (visual or auditory). Similarly, in a study featuring a reaction task in a virtual environment in which a ball hits a wall, trimodal cues resulted in faster responses than bimodal cues, which in turn resulted in faster responses as compared to unimodal cues (Sella, Reiner, & Pratt, 2013). These studies indicate that sensory integration can accelerate cortical activity, resulting in faster responses and less errors. However, limitations concerning sensory integration have been found as well. In a study by Abrahamse and colleagues (Abrahamse, Van der Lubbe, & Verwey, 2009) participants performed a serial reaction time task to study sequence learning using four different stimulus positions that were presented either unimodally (visual or tactile) or bimodally (visual-tactile). It was found that performance in the visual-only and visual-tactile condition was similar, showing no benefit in performance of the addition of tactile stimuli. Based on their findings it was concluded that awareness, learning and performance of sequence learning was not aided by sensory integration effects. Furthermore, Santangelo, Van der Lubbe, Belardinelli & Postma (2008) studied sensory integration of spatially non-predictive bimodal cues on the exogenous orienting of spatial attention. In a visual target detection task, spatially non-predictive visual, auditory, or bimodal cues were presented either left or right of a fixation cross. Although a superadditive effect of bimodal cues was found in neural activity, no benefit of bimodal cues was observed in the cuing effect. Their findings suggest that the presence of sensory integration is not necessarily reflected in behavioral effects, at least not in tasks with a low perceptual demand.. This does not rule out that sensory integration has no behavioral effects on tasks of a certain cognitive complexity. An example is the integration of incongruent bimodal (audiovisual) speech patterns. The visual image of a person speaking the syllable “fah” will alter the perception of the auditorily spoken syllable “gah”, resulting in the illusion of hearing the syllable “fah”. This bimodal speech integration effect is known as the McGurk effect (MacDonald & McGurk, 1978).

A task often used to study sequences of stimuli is the discrete sequence production task. In the DSP task a stimuli and responses alternate; once a response is provided, a new stimulus is presented. Implicit sequence learning can be studied by repeating certain sequential stimulus patterns (Abrahamse, Van der Lubbe, & Verwey, 2009). However, in the DSP task motor preparation and execution may occur in parallel, making it difficult to study the preparatory process in isolation. To separate motor preparation from execution, the DSP task was modified based on the precuing paradigm by Rosenbaum (1980). In the precuing paradigm a precue containing information regarding the expected response is followed by a brief preparation interval and a Go/Nogo signal. During this preparation interval the participants to were able to prepare their motor response without actually executing it. This modified sequence production task has been used to study the effect of sequence familiarity on general motor preparation and visual working memory (De Kleine & Van der Lubbe, 2011). In the present study we employed a similar sequence production task to study if a bimodal format of stimulus presentation resulted in improved performance and improved task learning. In order to prevent sequence learning effects, only unfamiliar sequences were used. Furthermore, the use of multiple sessions allowed us to study if a task-specific learning effect was present and if the bimodal condition affected this learning effect.

Event related potentials (ERPs) have been used when studying movement preparation and in this study we used the contingent negative variation (CNV) to measure the processes that precede motor execution of the presented sequence (Leuthold and Jentzsch, 2002). The CNV is a negative going wave that can be observed during the interval between a warning stimulus and a Go/Nogo signal that signifies execution. It seems that the CNV does not reflect a single process, as different studies have found various aspects regardin the CNV. The central CNV has been linked to general motor preparation according by De Kleine and Van der Lubbe (2011), who found a decreased central CNV for familiar sequences as compared to unfamiliar sequences. They concluded that familiar sequences resulted in a lower load on general motor preparation. However, response complexity has also found to be reflected in the central CNV (Cui et al., 2000). In their study, participants had to perform motor tasks of various complexity. Participants opposed their thumb and index finger three times (simple motor task), or alternated the thumb opposition with the little finger during the second opposition (complex motor task). An increased CNV was found for the complex task, suggesting that response complexity influenced the central CNV. The number of prepared responses has been linked to the central CNV as well (Schröter, & Leuthold, 2009). They found in increased CNV when preparing three-key responses as compared to single-key responses, suggesting that the number of preprogrammed motor responses was reflected in the central CNV. Overall, it seems that previous studies have looked at different aspects that influence general motor preprogramming. The parietal CNV has been linked to visual-spatial processes (Zimmer, 2008).

In the present study, we examined the differences in performance between unimodal and bimodal formats of stimulus presentation in a movement sequence task and related these to the CNV measure derived from the EEG. We predicted that if a bimodal format of stimulus presentation benefits from the use of both the visual and auditory working memory and/or the sensory integration mechanism, this would result in better performance than a unimodal format of stimulus presentation. In other words, we would expect a lower RT and a higher PC in the bimodal condition as compared to the unimodal (visual-only) condition. Regarding our expectations of the CNV, the present study only used unfamiliar sequences and response complexity was similar for all sequences. Any changes in the central CNV were likely to be linked to the amount of motor preprogramming. Thus, more motor preprogramming would result in an increased CNV. If we assume that a bimodal advantage results in better retention of sequence information in working memory and increased preparation, we would to see increased negativity in the central area. Furthermore, if participants became better after practice, motor preprogramming would increase and result in an increase of the central CNV. Lastly, the improved performance in the bimodal condition as compared to the unimodal visual condition would lead to a further increase of performance after practice and result in an increased CNV as well.

Methods

Participants

Fourteen students (four men, ten women), aged 19-37 years (mean: 22 years), from the University of Twente participated in this study. All participants were right-handed as assessed with the Annett Handedness Inventory (Annett, 1970) and reported normal or to corrected-to-normal vision. The participants signed an informed consent form before the experiment started. Twelve participants received credits for their participation, the remaining two participated on a voluntary basis. The local ethics committee of the Faculty of Behavioral Sciences of the University of Twente approved the study.

Stimuli and task

Participants placed the index finger, middle finger, ring finger and little finger of their right hand on the j, k, l, ; keys. In each trial four visual, auditory or bimodal stimuli were presented. In the case of a subsequent Go signal four spatially corresponding keypresses were to be executed by the participant. Figure 1 shows the presentation of the visual stimuli during a trial. At the start of the trial a default screen was shown consisting of a fixation cross (1.3°) with four horizontally aligned squares (2.5°) in the lower field, subtending a total visual angle of (12.5°). The screen background was black and the fixation cross and squares were drawn in grey. This screen was shown for 1000 ms. Then a sequence consisting of four stimuli (either visual, auditory or bimodal) was presented. Visual stimuli were presented by filling one of the four squares yellow for 500 ms. Stimulus onset asynchrony was 500 ms. Auditory stimuli were presented by playing one of four different tones through the loudspeakers for 500 ms, also with a stimulus onset asynchrony of 500 ms. Tones were selected by using four notes of equal tonal distance from the major scale of G. This resulted in the tones G4 (392.0 Hz), B4, (493.9 Hz), D5 (587.3 Hz) and F#5 (740.0 Hz). Volume levels of both stereo channels were adjusted to add a spatial component to each tone so that from left to right in the stereo field, the tones were presented in ascending order. Tones G4 and F#5 were presented through the left and right speaker only, respectively. For B4 the volume levels of the left and right stereo channels were set to 80% and 32.5% of the maximum volume. For D5 these values were reversed, 32.5% for the left and 80% for the right channel. These volume settings resulted in the remaining two auditory stimulus locations (left of center and right of center, respectively). Participants all reported being able to discriminate between the four auditory stimulus locations. Bimodal stimuli consisted of the simultaneous presentation of the visual and corresponding auditory stimulus. After the sequence was presented, the default screen was shown for 1500 ms. Participants were instructed to prepare to enter the presented sequence upon presentation of a Go signal. If a participant pressed a key during the preparation interval, a feedback screen containing the words “Too early” was shown for 1000 ms. In case of a Go trial, the color of the fixation cross turned blue for 100 ms and participants had to press the keyboard keys corresponding to the presented sequence of yellow squares (in the visual and bimodal condition) and/or auditory tones (in the auditory and bimodal condition). Participants were requested to enter the presented sequence as fast and accurately as possible. Each keypress resulted in the appropriate square turning yellow for 100 ms (visual and bimodal condition) and/or the appropriate tone playing for 100 ms (auditory and bimodal condition). After entering the sequence correctly, a feedback screen was presented showing “Correct” for 1000 ms. If a participant had made one or more incorrect keypresses, each incorrectly pressed key was shown on screen for 1000 ms. For example, if the first key of the sequence was pressed incorrectly, the feedback screen would show “Key 1 incorrect” for 1000 ms. If more than 2000 ms had passed between two keypresses before the complete sequence was entered ,“Too late” was shown for 2000 ms. In a Nogo trial the fixation cross turned red for 2000 ms. Any keypress during this interval resulted in the feedback screen containing the words “Do not press a key” for 2000 ms. Each trial ended with the screen turning blank for 500 ms.

Sequence generation

Sequences were generated with the limitation that keys were not repeated on consecutive elements in the sequence (e.g. 1334; numbers from 1 to 4 are denoting stimulus locations from left to right) or with one other element in between (e.g. 1232). This was done to prevent finger-specific effects. Based on these constraints 48 different sequences were generated for this experiment (4 x 3 x 2 x 2).

Procedure

Before starting the experiment, participants first ran a tutorial explaining the task. In this tutorial examples of sequences were given for each modality and participants could try out the keys on the keyboard to see and hear the feedback. Lastly, ten trials were practiced in each modality to familiarize the participant with the trial structure and the tones in the auditory and bimodal conditions. No behavioral data was registered during the tutorial. After preparing the EEG setup and if the participant had no further questions, the experiment started.

The experiment contained two sessions of 312 trials each. Each session consisted of three blocks of 104 trials. All trials in a block were presented in the same modality. A block consisted of 96 Go trials (92%) and 8 Nogo trials (8%). In a Go trial one of the 48 sequences was randomly chosen. Each sequence occurred exactly twice as a Go trial in each block. In a Nogo trial one of the 48 sequences was randomly chosen. Go and Nogo trials were randomly intermixed. Block order in the first session (for example visual, auditory, bimodal) was the same in the second session and was counterbalanced across participants to prevent order specific effects. Between blocks was a one minute break. Between both sessions was a five minute break.

Apparatus and data acquisition

Each participant performed the experiment on a Pentium 4 computer with a QWERTY keyboard. E-prime version 2.0 was used for the presentation of stimuli, recording of keyboard responses, and the external communication of trigger data with a Quick-Amp EEG data amplifier through the parallel port. A 17 inch CRT monitor screen with a refresh frequency of 75 Hz was used for task presentation. For sound output two loudspeakers were connected to the computer. The loudspeakers were placed in front of the screen, 93 cm apart, on each side of the keyboard. Output levels were set so that the sound output had a sound level of 50 dB. EEG data was recorded using Brain Vision Recorder version 1.05. An actiCAP electrode cap with 64 active electrodes was used to record brain activity, using the extended 10/20 system. The ground (GND) electrode was placed on the forehead and the reference (REF) electrode was placed on the left mastoid. EOG was recorded bipolarly using Ag/AgCl ring electrodes, with electrodes above and below the left eye for recording the vertical EOG and electrodes at the outer canthus of each eye for recording the horizontal EOG. Electrode impedance was kept below 10 kΩ for all electrodes. EEG and EOG data were sampled at 1000 Hz.

Data analysis

Due to a high number of errors (more than 50% in the auditory condition), one participant was excluded from data analysis. Due to procedural errors, EEG data of another participant was excluded from data analysis. The first 500 ms of the preparation interval were not used in the analysis in order to prevent confounding effects due to arousal from the auditory stimuli. For EEG data analysis the 1000 ms interval of each trial prior to the Go/Nogo signal was used. The -1100 – -1000 ms interval was used as a baseline. Trials containing horizontal eye movements were removed by using an hEOG amplitude threshold of 60 µV (positive direction) and -60 µV (negative direction) with a time tolerance of 100 ms. Trials were checked for artifacts (maximal allowed voltage step of 120 µV/ms and values larger than +/- 200 µV). Electrode channels with an excessive number of artifacts were replaced by using the averaged values of the neighboring channels. In all other cases segments containing artifacts were discarded. The Gratton & Coles ocular correction procedure (Gratton, Coles, & Donchin, 1983) was used to correct the EEG for EOG artifacts. Finally, all segments in every modality condition were averaged for each participant and were filtered with a 16 Hz low-pass filter. Grand averages were constructed from these averages and were used for topographical mapping of cortical activity as well as the generation of ERP graphs.

Behavioral data. The response time (RT) was defined as the time between onset of the Go-signal and pressing the first key of the sequence on the keyboard, and as the time between two consecutive key presses while entering a sequence. Trials excluded from RT analysis were trials containing errors as well as the first two trials of every block. Furthermore, trials in which the sum of all RTs (total RT) in one sequence deviated more than 3 SD from the overall mean total RT in that block were also excluded from analysis (De Kleine & Van der Lubbe, 2011; De Kleine & Verwey, 2009). This procedure led to the removal of 1.5% of the trials. The Percentage Correct (PC) for each block was calculated as the percentage of correct keypresses (De Kleine & Van der Lubbe, 2011) in all trials that required a response. An arcsine transformation was applied to the PC data in order to make the distribution normal (Abrahamse & Verwey, 2008). An analysis of variance (ANOVA) with repeated measures was performed for statistical evaluation of the mean RT and mean arcsine-transformed PC, using stimulus Modality (3), Session (2) and Key (4) as within subjects factors.

EEG data. Topographical maps of different time intervals for the preparation interval of the three modalities for both sessions were used for visual inspection and comparison. A repeated measures ANOVA was performed on the CNV for ten 100 ms intervals prior to the Go-Nogo signal using the factors Anterior-Posterior axis (5), Lateral-Medial axis (7), Modality (3) and Session (2). The five levels for the factor Anterior-Posterior were Frontal (electrodes F5, F3, F1, Fz, F2, F4, and F6), Fronto-central (FC5, FC3, FC1, FCz, FC2, FC4, and FC6), Central (C5, C3, C1, Cz, C2, C4, and C6), Centro-parietal (CP5, CP3, CP1, CPz, CP2, CP4, and CP6), and Parietal (P5, P3, P1, Pz, P2, P4, and P6). The seven levels for the factor Lateral-Medial were Left-lateral (F5, FC5, C5, CP5, and P5), Left-middle (F3, FC3, C3, CP3, and P3), Left-medial (F1, FC1, C1, CP1, and P1), Central (Fz, FCz, Cz, CPz, and Pz), Right-medial (F2, FC2, C2, CP2, and P2), Right-middle (F4, FC4, C4, CP4, and P4), and Right-lateral (F6, FC6, C6, CP6, and P6). Effects were only considered significant when two or more consecutive intervals showed an effect (Talsma, Wijers, Klaver, & Mulder, 2001). The means of main effects and interaction effects were further investigated. Greenhouse-

Geisser epsilon correction was applied for the degrees of freedom whenever the assumption of sphericity was violated.

Results

Behavioral measures RTs and Percentage Correct (PC) as a function of Session and Modality are compiled in Table 1.

Response times. There was a main effect of Modality on response time, $F(2, 24) = 19.77$, $\varepsilon = 0.66$, $p < .001$, $\eta_p^2 = 0.62$. Post-hoc contrast analysis showed that response times in the visual condition were lower than in the auditory condition (368 vs. 466 ms), $F(1, 12) = 22.94$, $p < .001$, $\eta_p^2 = 0.66$. Also, response times in the bimodal condition were lower than in the auditory condition (375 vs. 466 ms), $F(1, 12) = 20.67$, $p = .001$, $\eta_p^2 = 0.63$. Response times in the visual and bimodal condition did not differ significantly. A main effect of Session on response times was found, $F(1, 12) = 57.60$, $p < .001$, $\eta_p^2 = 0.83$, indicating that participants became faster in the second session (425 vs. 380 ms). Participants were faster on later keypresses (528, 374, 363, 345 ms respectively for key 1-4), as shown by the effect of Key, $F(3, 36) = 45.51$, $\varepsilon = 0.42$, $p < .001$, $\eta_p^2 = 0.79$, see Figure 2. An interaction between Modality and Key was found, $F(6, 72) = 7.41$, $\varepsilon = 0.42$, $p = .001$, $\eta_p^2 = 0.38$. Post-Hoc contrast analysis showed that the RT difference between the first and second keypress was greater in the visual condition than in the auditory condition (197 vs. 113 ms), $F(1, 12) = 27.64$, $p < .001$, $\eta_p^2 = 0.70$. The same was found the between the bimodal condition and the auditory condition (152 vs. 113 ms), $F(1, 12) = 6.32$, $p = .027$, $\eta_p^2 = 0.35$. Furthermore, the bimodal and visual condition differed as well (152 vs. 197 ms), $F(1, 12) = 25.02$, $p < .001$, $\eta_p^2 = 0.68$. While participants responded faster on the first key in the bimodal condition (bimodal 498 vs. visual 521 ms), the second key was pressed faster in the visual condition (bimodal 346 ms vs. visual 324 ms), see Figure 2. This observation could be explained by an arousal effect due to the auditory stimuli, resulting in a faster initial reaction in the bimodal condition.

Percentage correct. The results from the repeated measures ANOVA in this section are from the arcsin transformed PC data. A main effect of Modality was found on PC, $F(2, 24) = 46.86$, $\varepsilon = 0.31$, $p < .001$, $\eta_p^2 = 0.80$. Post-hoc contrast analysis showed that subjects made less errors in the visual compared to the auditory condition (94.7 vs. 72.6%), $F(1, 12) = 43.24, p < .001, \eta_p^2 = 0.78$. Similarly, less errors were made in the bimodal condition compared to the auditory condition (95.0 vs. 72.6%), $F(1, 12) = 72.46$ $p < .001, \eta_p^2 = 0.86$. The PC between the visual and bimodal condition did not differ significantly, showing no advantage of bimodally presented stimuli (see Figure 2). Less errors were made in the second session, as shown by the main effect of Session (86.7 vs. 88.1%), $F(1, 12) = 8.09$, $p = .015, \eta_p^2 = 0.40$. An effect of Key was found, $F(3, 36) = 27.92, p < .001, \eta_p^2 = 0.70$, as more errors were made on subsequent keypresses (90.6%, 88.0%, 85.9%, 85.0% repectively for key 1-4), see Figure 2. No interaction effects were found.

EEG analysis. Figure 3 displays topographic maps with cortical activity during five 200 ms intervals prior to the Go/Nogo signal, for each modality and session. Visual inspection shows an increase in positivity in the anterior region as well as an increase in negativity in the posterior region over the course of the preparation interval, for all modalities. There seems to be a difference in cortical activity between modalities, in particular between the auditory condition and the visual/bimodal condition. Furthermore, cortical activity seems to differ between sessions for all modalities. Table 2 depicts the results of the repeated measures ANOVA for the ten 100 ms intervals during the preparation period. The results from the ANOVA provided in this section are from the latest interval that reached significance. A main effect of Anterior-Posterior axis was found for all but the first 100 ms interval, due to positivity in the fronto and frontocentral regions and negativity in the central, centroparietal and parietal regions, $F(4, 44) = 35.27, \varepsilon = 0.41, p < .001, \eta_p^2 = 0.76$. On average, the mean cortical activity was lower in the visual condition (-0.55 μV) compared to the bimodal condition (-0.43 μV), which in turn was lower than the auditory condition (-0.10 μV), as shown by the main effect of Modality for the last 300 ms of the preparation interval, $F(2, 22) = 5.40, p = .012, \eta_p^2 = 0.33$. The main effect of Session shows that there was a stronger negativity in the second session (-0.51 μV) than in the first (-0.21), $F(1, 11) = 5.53, p = .038, \eta_p^2 = 0.33$. An interaction between Anterior-Posterior axis and Lateral-Medial axis was found, $F(24, 264) = 1.64, p$

= .034, $\eta_p^2 = 0.13$, with the strongest positivity and negativity located at F3 (2.72 μV) and P1 (-2.76 μV), respectively. There was an interaction between Lateral-Medial axis and Modality, *F*(12, 132) = 2,58, $p = .004$, $\eta_p^2 = 0.19$, showing that while the visual and bimodal modalities showed a strongest negativity at one location (right medial axis, the auditory modality showed the strongest negativity at two locations, namely the left and right medial axes). An interaction between Modality and Session was found, *F*(2, 22) = 4.56, $p = .022$, $\eta_p^2 = 0.29$, as the visual and auditory modality showed an increase in negativity between sessions, while the bimodal modality showed a no difference or in some intervals even a decrease in negativity. Finally, a three-way interaction between Anterior-Posterior axis, Lateral-Medial axis and Session was found, *F*(24, 264) = 1.61, $p = .039$, $\eta_p^2 = 0.13$. The greatest change in voltage between sessions was found to be at C1 (an increased negativity of 1.43 μV in the second session). See Figure 4 for the corresponding ERP.

Discussion

In this study we examined if a bimodal format of stimulus presentation in a sequence movement task would lead to improved performance and if it would influence task-specific learning as compared to unimodal stimuli, as a result of improved working memory usage and/or sensory integration. We predicted that a beneficial effect of bimodal stimuli would result in an increased CNV, as well as an effect of practice on the CNV for all modalities. We used behavioral data (response times and percentage correct) to determine performance and to see if a task-specific learning effect was present across sessions. The CNV data recorded during the preparation interval between stimulus presentation and the go/nogo signal was analyzed to see which cortical areas may have been related to the differences in performance between modalities and across sessions.

Behavioral results showed that performance in the visual and bimodal conditions did not differ for both RT and PC, indicating that bimodally presented stimuli did not lead to improved performance compared to the unimodal visual condition. Performance in the auditory condition was significantly worse for both RT and PC compared to the visual and bimodal condition, showing that the task with only auditory stimuli was more challenging. For all modalities a decrease in RT as well an increase in PC was found in the second session for all modalities, showing that the participants became faster and made less errors after practice. As no interaction effect was found for modality and session, there does not seem to be a beneficial effect of bimodal stimuli on task-specific learning.

The results from the EEG analysis indicate a general effect of modality on the CNV. The absence of a difference between the visual and bimodal condition in the contrast analysis seems to coincide with the finding that no difference was found between these conditions in the behavioral analysis as well. On the other hand, a difference between the auditory and the visual/bimodal conditions was found in both the EEG and behavioral data. Since the effect of modality in the EEG analysis was not related for a specific cortical area, care has to be taken when interpreting this result. At the very least, it seems that the differences in behavioral data due to different stimulus modalities are reflected in the EEG on a general level. Since different preparatory processes are reflected by topographic differences in CNV amplitude (Leynes, Allen, & Marsh, 1998), it is not possible to attribute either a particular process or a combination of processes to the modality-specific findings in the present study.

We observed an effect of Session at the contralateral motor cortex. At this site the negative CNV increased for all modalities in the second session. If we assume that participants improve task-specific skills as a result of practice, then the change in CNV at this area may reflect a change in motor preprogramming. Response times as well as percentage correct improved for all modalities in the second session, suggesting that working memory performance improved. In relation to our study, this would imply that cognitive demands on working memory may have been lower in the second session for all three modalitie, resulting in improved general motor preparation. This corresponds with the behavioral data that response times were lower in the second session for all three modalities. To the best of our knowledge, this is the first study to show an effect of task-specific learning in the central CNV.

A location-unspecific interaction between stimulus modality and session was found. This interaction showed that the effect of session was different for the bimodal condition. A reserved interpretation has to be made regarding this finding, but it seems that the effect of practice had a different effect on the bimodal CNV as compared to the visual CNV. It may be possible that the task-specific bimodal learning effects may require more practice before they establish clearly in the CNV. Further study may provide insights in the development of the bimodal performance and CNV after extensive practice.

As we found in this study, auditory stimulus presentation resulted in inferior performance compared to a visual or bimodal format of stimulus presentation. It may have been possible that auditory stimuli required additional cognitive processing to build the internal representation of the sequence that was used. However, a specific effect of modality on the parietal CNV was not found, suggesting that visual-spatial processes were similar for all modalities. Performing this study with different types of distractor tasks during the preparation interval could provide insight in the type of memory system used for storing auditory sequences. Furthermore, auditory memory may be better suited for storing rhythmic information rather than pitch information. Performing a sequence production task using sequences featuring temporal patterns rather than spatial locations may result in a better invocation of auditory memory and may therefore be beneficial in a bimodal format of sequence presentation. For example, the flash fission illusion is an illusion where the number of auditory beeps influences the number of perceived visual flashes, showing that the auditory system is dominant in processing temporal patterns (Innes-Brown, & Crewther, 2009).

Although sensory integration seems to be an involuntary process, attentional control seems to be a factor in the degree of sensory integration (Talsma & Woldorff, 2005). In their study a rapid serial visual presentation stream of letters was used along with auditory and visual objects that had to be responded to. Attending both stimulus modalities resulted in better performance (lower RT and higher PC) as compared to attending only one stimulus modality during audiovisual stimulus presentation, suggesting that attentional control seems to be a factor in the degree of sensory integration. Due to visual dominance, participants may have subconsciously chosen to attend the visual modality only during bimodal presentation in the present study, thereby decreasing the effectiveness of sensory integration.

While in this study no beneficial effect of bimodal sequences was found, bimodal sequences may be usable for sequence learning. For example, Cohen, Evans, Horowitz and Wolfe (2011) found that when comparing musicians with nonmusicians, musicians excelled on an auditory memory task, but did not perform better in a visual memory task. Similarly, it may be possible that musicians perform differently than nonmusicians on a bimodal sequence production task.

To conclude, the present study found no positive effect on working memory performance for bimodally (audiovisual) as compared to unimodally (visual) presented movement sequences. From the EEG analysis, it seemed that task-specific learning was reflected in the contralateral left-medial CNV. This finding may be related to improved general motor preparation as a result of better working memory usage due to practice. There are general indicators that preparation in the bimodal condition develops differently compared to the visual condition after practise, but further research would be required to get a clearer view on this.

Table 1

Mean RTs (in ms) and PC (in %) as a function of Session for each of the three modality conditions

	RT			PC		
Session	Visual	Auditory	Bimodal	Visual	Auditory	Bimodal
1	391	489	397	94.2 (77.3)	70.9 (58.0)	94.9 (78.1)
2	346	443	353	95.1 (78.7)	74.2 (60.4)	95.0 (78.5)

Note. Listed in parentheses are the means of the arcsin transformed PC data used in the ANOVA analysis.

Table 2

Intervals in which significant effects were found in the repeated measures ANOVA

Factor(s)	Highest / Lowest F-statistic	Significant interval(s)	Pattern and means (μV) in the last significant interval
AnPos *F*(4, 44)	35.27 8.42	-900 – 0	Highest / Lowest value F (2.21) P (-2.03)
Modality *F*(2, 22)	5.40 4.24	-300 – 0	Visual (-0.55) Bimodal (-0.43) Auditory (-0.10)
Session *F*(1, 11)	12.85 5.53	-600 – -400 -300 – 0	Session 1 (-0.21) Session 2 (-0.51)
AnPos * LatMed *F*(24, 264)	4.40 1.64	-1000 – -100	Highest / Lowest value F3 (2.72) P1 (-2.76)
LatMed * Modality *F*(12, 132)	2.58 1.89	-600 – 0	Visual (-1.52) and Bimodal (-1.11) strongest negativity at left-medial axis, Auditory (-0.46 and -0.52) strongest negativity at left and right-medial axis.
Modality * Session *F(2, 22)*	5.23 4.34	-400 – -100	Increase in negativity: Visual (-0.35) Auditory (-0.50) Bimodal (-0.05)
AnPos * LatMed * Session *F*(24, 264)	1.62 1.61	-200 – 0	Strongest decrease in negativity between sessions C1 (1.43)

Note. Interval values are in milliseconds, relative to the Go/Nogo signal. Factors used in the repeated measure analysis are Anterior-Posterior axis (5), Lateral-Medial axis (7), Modality (3) and Session (2). Degrees of freedom of each F-statistic and its error term are given, prior to an eventual correction for a violation of sphericity. Only effects that were statistically significant in at least two consecutive intervals are listed.

AnPos = Anterior-Posterior; LatMed = Lateral-Medial

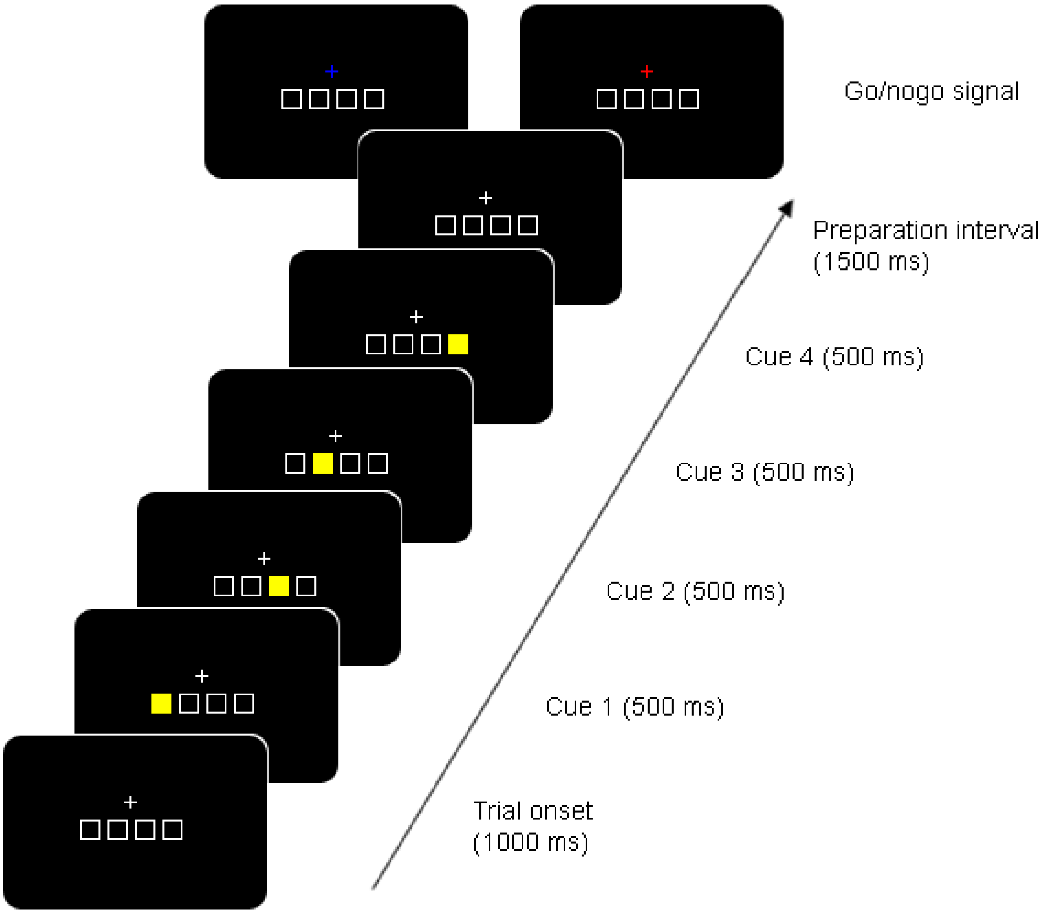

Figure 1. Presentation of a trial in the visual condition. In the auditory condition the trial onset screen remained until the go/nogo signal was presented while for each cue the appropriate auditory stimulus was played. The bimodal condition featured stimulus presentation in both modalities.

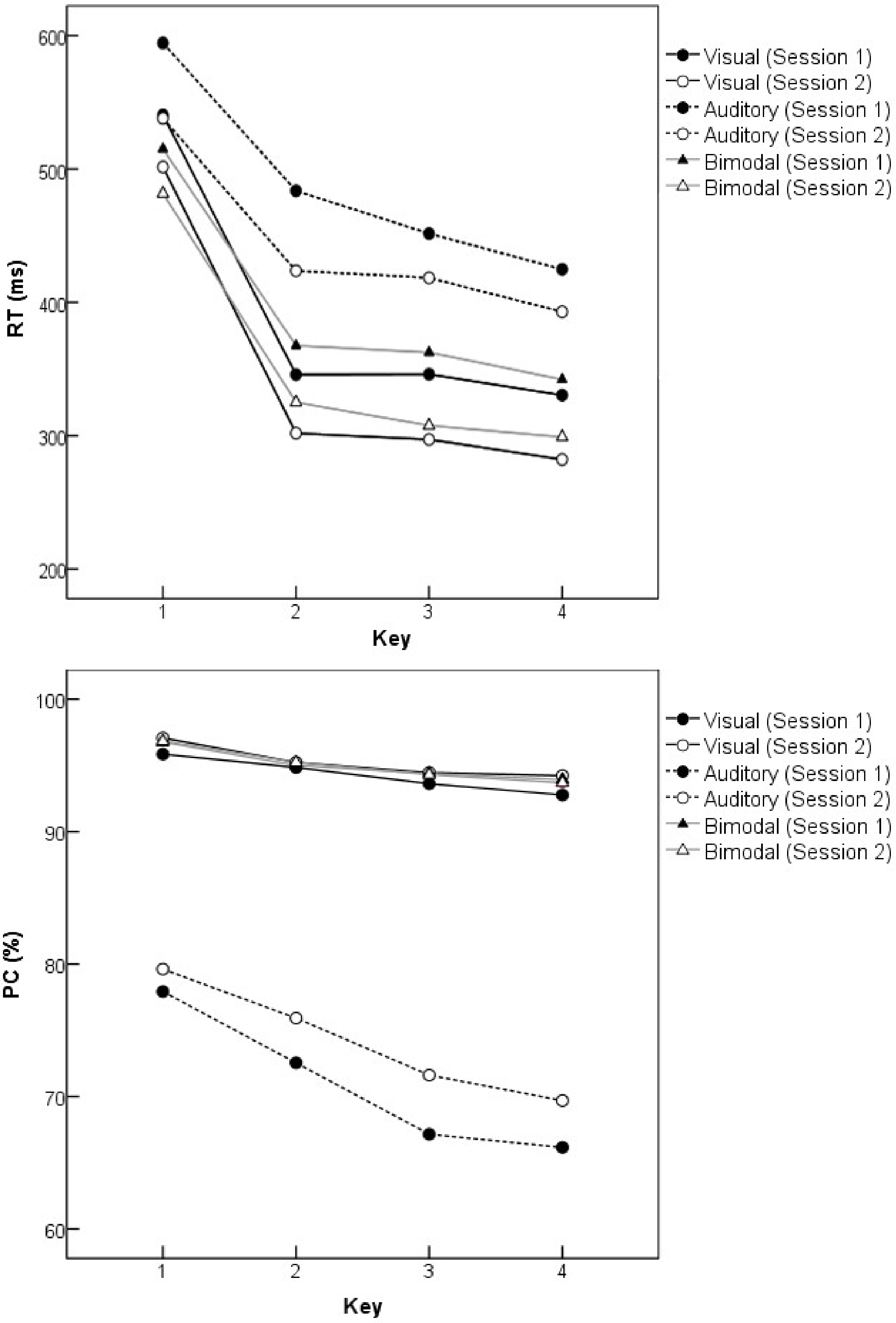

Figure 2. Mean response time (RT) and percentage correct (PC) as a function of Key, Modality and Session.

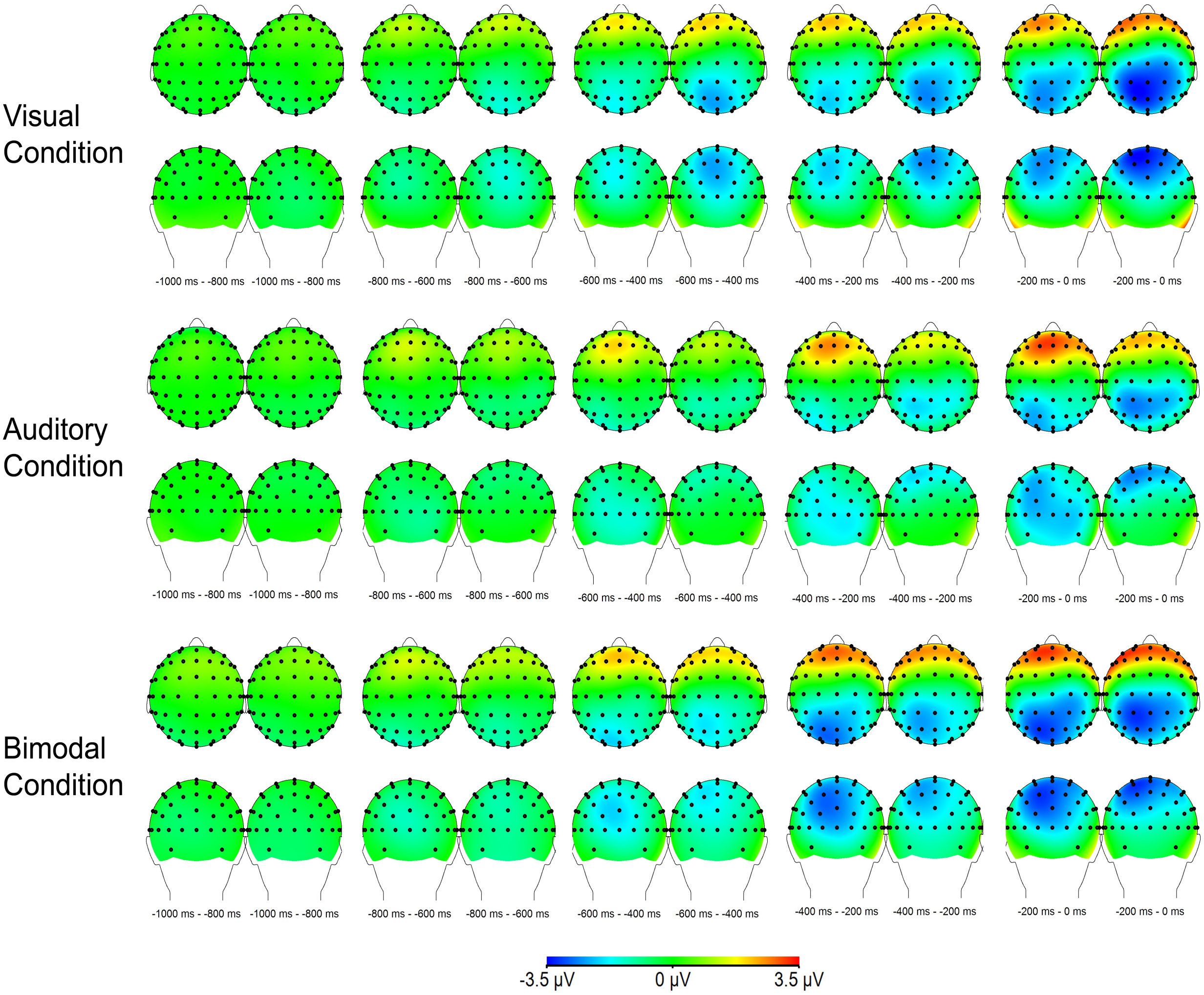

Figure 3. Topographic maps depicting the top and rear views of the averaged activity for each of the three modality conditions. Each column shows a 200 ms interval of the response preparation, relative to the Go/Nogo signal (0 ms). Of each pair, the left map shows the activity in the first session, and the right maps shows the activity in the second session.

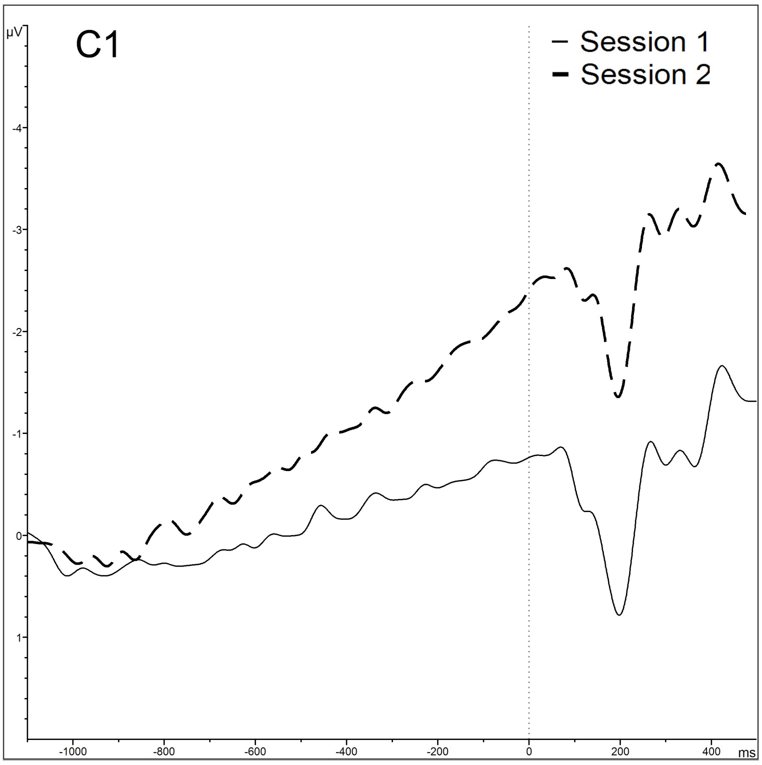

Figure 4. Event-related brain potential (ERP) at C1 and for the factor Session.

References

Abrahamse, E. L., van der Lubbe, R. H., & Verwey, W. B. (2009). Sensory information in perceptual-motor sequence learning: visual and/or tactile stimuli. *Experimental Brain Research, 197*, 175-183. doi:10.1007/s00221-009-1903-5

Abrahamse, E. L., & Verwey, W. B. (2008). Context dependent learning in the serial RT task. *Psychological research, 72*, 397-404. doi:10.1007/s00426-007-0123-5

Annett, M. (1970). A classification of hand preference by association analysis. *British Journal of Psychology, 61,* 303-321. doi:10.1111/j.2044-8295.1970.tb01248.x

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation, 8*, 47-89. doi:10.1016/S0079-7421(08)60452-1

Brunia, C. H. M. (2004). Slow potentials in anticipatory behavior. *Journal of Psychophysiology, 18*, 59-60. doi:10.1027/0269-8803.18.23.59

Cui, R. Q., Egkher, A., Huter, D., Lang, W., Lindinger, G., & Deecke, L. (2000). High resolution spatiotemporal analysis of the contingent negative variation in simple or complex motor tasks and a non-motor task. *Clinical Neurophysiology, 111*, 1847-1859. doi:10.1016/S1388-2457(00)00388-6

Cohen, M. A., Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). Auditory and visual memory in musicians and nonmusicians. *Psychonomic bulletin & review, 18*, 586-591. doi:10.3758/s13423-011-0074-0

De Kleine, E., Van der Lubbe, R. H. J. (2011). Decreased load on general motor preparation and visual-working memory while preparing familiar as compared to unfamiliar movement sequences. *Brain and Cognition, 75,* 126-134. doi:10.1016/j.bandc.2010.10.013

De Kleine, E., & Verwey, W. B. (2009). Representations underlying skill in the discrete sequence production task: Effect of hand used and hand position. *Psychological Research, 73,* 685-694. doi:10.1007/s00426-008-0174-2

Fischer, M. H. (2001). Probing spatial working memory with the Corsi blocks task. *Brain and Cognition, 45*, 143-154. doi:10.1006/brcg.2000.1221

Goolkasian, P., & Foos, P. W. (2005). Bimodal format effects in working memory. *The American Journal of Psychology, 118,* 61-78.

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for the off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*, 468-484. doi:10.1016/0013-4694(83)90135-9

Innes-Brown, H., & Crewther, D. (2009). The impact of spatial incongruence on an auditory-visual illusion. *PloS one, 4*, e6450. doi:10.1371/journal.pone.0006450

Leuthold, H., & Jentzsch, I. (2002). Distinguishing neural sources of movement preparation and execution: An electrophysiological analysis. *Biological psychology*, 60(2), 173-198. doi:10.1016/S0301-0511(02)00032-7

Leynes, P. A., Allen, J. D., & Marsh, R. L. (1998). Topographic differences in CNV amplitude reflect different preparatory processes. International Journal of Psychophysiology, 31, 33-44. doi:10.1016/S0167-8760(98)00032-4

Mastroberardino, S., Santangelo, V., Botta, F., Marucci, F. S., & Belardinelli, M. O. (2008). How the bimodal format of presentation affects working memory: an overview. *Cognitive processing, 9*, 69-76. doi:10.1007/s10339-007-0195-6

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24*, 253-257. doi:10.3758/BF03206096

Monfort, V., & Pouthas, V. (2003). Effects of working memory demands on frontal slow waves in time-interval reproduction tasks in humans. *Neuroscience letters, 343*, 195-199. doi:10.1016/S0304-3940(03)00385-9

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive brain research, 14*, 115-128. doi:10.1016/S0926-6410(02)00066-6

Rhodes, B. J., Bullock, D., Verwey, W. B., Averbeck, B. B., & Page, M. (2004). Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives. *Human movement science, 23,* 699-746. doi:10.1016/j.humov.2004.10.008

Rosenbaum, D. A. (1980). Human movement initiation: specification of arm, direction, and extent. *Journal of Experimental Psychology: General, 109*, 444. doi:10.1037/0096-3445.109.4.444

Santangelo, V., Mastroberardino, S., Botta, F., Marucci, F. S., & Belardinelli, M. O. (2006). On the influence of audio-visual interactions on working memory performance: a study with non-semantic stimuli. *Cognitive Processing, 7,* 187-187. doi:10.1007/s10339-006-0136-9

Santangelo, V., Van der Lubbe, R. H., Belardinelli, M. O., & Postma, A. (2008). Multisensory integration affects ERP components elicited by exogenous cues. *Experimental Brain Research, 185*, 269-277. doi:10.1007/s00221-007-1151-5

Schröter, H., & Leuthold, H. (2009). Motor programming of rapid finger sequences: Inferences from movement-related brain potentials. *Psychophysiology, 46*, 388-401.

doi:10.1111/j.1469-8986.2008.00772.x

Sella, I., Reiner, M., & Pratt, H. (2013). Natural stimuli from three coherent modalities enhance behavioral responses and electrophysiological cortical activity in humans. *International Journal of Psychophysiology. 93,* 45-55. doi:10.1016/j.ijpsycho.2013.11.003

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *Journal of cognitive neuroscience, 17,* 1098-1114. doi:10.1162/0898929054475172

Zimmer, H. D. (2008). Visual and spatial working memory: from boxes to networks. *Neuroscience & Biobehavioral Reviews, 32*, 1373-1395. doi:10.1016/j.neubiorev.2008.05.016