



DALHOUSIE UNIVERSITY

Retrieved from DalSpace, the institutional repository of
Dalhousie University

<https://dalspace.library.dal.ca/handle/10222/75713>

Version: Post-print

Publisher's version: Hassall, C. D., Quinlan, C. K., Turk, D. J., Taylor, T. L., & Krigolson, O. E. (2016). A preliminary investigation into the neural basis of the production effect. *Canadian Journal of Experimental Psychology*, 70, 139-146. doi: 10.1037/cep0000093

May 2019

This manuscript is a pre-publication version of an article published as:

Hassall, C. D., Quinlan, C. K., Turk, D. J., Taylor, T. L., & Krigolson, O. E. (2016). A preliminary investigation into the neural basis of the production effect. *Canadian Journal of Experimental Psychology*, 70, 139-146.

This posted article might not exactly replicate the final version published in the above-noted journal; it is not the copy of record. As part of the publication process, copyright was transferred to the journal.

Please visit the publisher website to find the final published version of this work:

<https://www.apa.org/pubs/journals/cep/>

A preliminary investigation into the neural basis of the production effect

Cameron D. Hassall¹, Chelsea K. Quinlan², David J. Turk³, Tracy L. Taylor², & Olave E.

Krigolson¹

1. Neuroeconomics Laboratory, University of Victoria, Victoria, BC, Canada
2. Department of Psychology and Neuroscience, Dalhousie University, Halifax, NS, Canada
3. School of Experimental Psychology, Bristol University, Bristol, United Kingdom

Address correspondence to:

Cameron D. Hassall
School of Exercise Science, Physical and Health Education
University of Victoria
PO Box 1700 STN CSC
Victoria, BC, Canada
V8W 2Y2
Phone: (250) 721 7843
Email: chassall@uvic.ca

Abstract

Items that are produced (e.g., read aloud) during encoding typically are better remembered than items that are not produced (e.g., read silently). This “production effect” has been explained by distinctiveness: Produced items have more distinct features than non-produced items, leading to enhanced retrieval. The goal of the current study was to use electroencephalography (EEG) to examine the neural basis of the production effect. During study, participants were presented with words that they were required to read silently, read aloud, or sing while EEG data were recorded. Subsequent memory performance was tested using a yes/no recognition test. Analysis focused on the event-related brain potentials (ERPs) evoked by the encoding instruction cue for each instruction condition. Our data revealed enhanced memory performance for produced items and a greater P300 ERP amplitude for instructions to sing or read aloud compared to instructions to read silently. Our results demonstrate that the amplitude of the P300 is modulated by at least one aspect of production, vocalization (singing/reading aloud relative to reading silently), and are consistent with the distinctiveness account of the production effect. The ERP methodology is a viable tool for investigating the production effect.

Key words: production effect, memory, distinctiveness, electroencephalography, P300

It has been demonstrated that memory performance is greater for produced items (e.g., items read aloud) than non-produced items (e.g., items read silently: MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). This “production effect” is usually explained using a distinctiveness account – items with more distinct elements are specially encoded and more likely to be retrieved from memory than items with fewer distinct elements (see Ozubko & MacLeod, 2010). For example, during encoding, the additional distinct elements of articulation (the motor movements associated with saying the word aloud) and audition (hearing the word aloud) make items read aloud more robust in memory than items read silently. During retrieval these additional distinct features (e.g., articulation and audition) make items read aloud easier to recall than items read silently. Supporting a distinctiveness account, the production effect disappears when unstudied foil items are read aloud (Ozubko & MacLeod, 2010). Further, the production effect is present but reduced when study items are presented in between-participant (Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010) or pure list designs (Ozubko & MacLeod, 2010; see Fawcett, 2013, for a meta-analysis).

Although read aloud and read silently are often used for the two comparison instruction conditions in the production paradigm, other forms of production have also been used. These include mouthing (Castel, Rhodes, & Friedman, 2013; Fawcett, Quinlan, & Taylor, 2012; Forrin, MacLeod, & Ozubko, 2012; MacLeod et al., 2010), spelling (Forrin et al., 2012), writing (Forrin et al., 2012), typing (Forrin et al., 2012; Richler, Palmeri, & Gauthier, 2013), whispering (Forrin et al., 2012), and singing (Quinlan & Taylor, 2013). With the exception of Quinlan and Taylor (2013), all of the above studies have found that reading aloud results in a greater production effect compared to other forms of production (e.g., mouthing, spelling, writing, typing, and whispering – in particular, see Forrin et al., 2012). However, Quinlan and Taylor (2013) reported

that singing resulted in a greater production effect than reading aloud. This enhanced production effect for singing was given as further evidence for the distinctiveness account of the production effect: Compared to reading silently, reading aloud is assumed to involve both articulation and audition, whereas singing is assumed to involve the dynamic elements of intensity, pitch, and/or timbre in addition to articulation and audition (Quinlan & Taylor, 2013; Roederer, 2008).

While there is a large body of behavioural evidence that supports a distinctiveness account for the production effect (e.g. Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010; Ozubko & MacLeod, 2010; Quinlan & Taylor, 2013; Icht, Mama, & Algom, 2014; although see Bodner & Taikh, 2012; Fawcett, 2013), the neural mechanisms that underpin this effect have thus far not been explored. The current inquiry used electroencephalography (EEG) to chart the neural basis for the production effect. In addition, we also examined the relation between evoked cortical responses during encoding and subsequent memory to further examine the distinctiveness account of the production effect.

The P300 component of the human event-related brain potential (ERP) is a positive deflection in ERP waveforms observed 300 ms to 600 ms after stimulus presentation with a scalp distribution over the midline electrodes (Fz, Cz, Pz; see Polich, 2007, for a review). Although there are at least two subcomponents of the P300 (P300a, P300b), here we will focus on the P300b (and simply call it the P300) subcomponent that tends to be larger over parietal areas of cortex and is thought to index cognitive processing (Comerchero & Polich, 1999; Donchin & Coles, 1988). The P300 has been associated with a variety of cognitive processes, including context updating (Donchin, 1981) and resource allocation (Donchin, Karis, Bashmore, Coles, & Gratton, 1986). Additionally, a general finding in memory paradigms utilizing the ERP technique is that the amplitude of the P300 during item encoding is positively correlated with

subsequent memory performance (Paller, Kutas, & Mayes, 1987; Rushby, Barry, & Johnstone, 2002) and strength of memory formed during encoding (Crites, Devine, Lozano, & Moreno, 1998; Patterson, Pratt, & Starr, 1991).

Of particular relevance here, the P300 component has also been associated with distinctiveness. For instance, Karis, Fabiani, and Donchin (1984) manipulated distinctiveness via the font size of words (small, medium, large) at item encoding and found that: (a) the amplitude of the P300 was larger for distinct words as opposed to non-distinct words, and (b) distinct words were more likely to be recalled at test. Fabiani and Donchin (1995) replicated these effects for physically and semantically distinct words. They found that the amplitude of the P300 was greater for distinct as opposed to non-distinct words, and that participants were more accurate at recalling distinct words than non-distinct words. Together, these findings (and others) suggest that the amplitude of the P300 component is sensitive to distinctiveness in addition to being predictive of subsequent memory performance (see Fabiani & Donchin, 1995; Fabiani, Karis, & Donchin, 1990; Kamp, Forester, Murphy, Brumback, & Donchin, 2012; Karis et al., 1984). Given that the amplitude P300 ERP component is sensitive to distinctiveness, this component offers a potential index to delineate the processes that underpin this aforementioned production effect.

In the present study, during an initial study phase, participants were shown a series of words that they either had to read silently, read aloud, or sing aloud while EEG data were recorded. After a brief delay, participants completed a yes/no recognition test in which they were required to distinguish between studied and unstudied items. In line with the distinctiveness account of the production effect (i.e., Quinlan & Taylor, 2013) and the relation between the P300 ERP component and distinctiveness (e.g., Fabiani & Donchin, 1995), here we predicted that

P300 amplitude would scale with increasing distinctiveness and that this relation would be mirrored in memory performance. More specifically, in line with Quinlan and Taylor (2013), we hypothesized that reading aloud involves more distinct items than reading silently, and singing in turn involves more distinct dimensions than reading aloud. Given the relation between P300 amplitude and distinctiveness, we predicted that the amplitude of the P300 evoked by an instruction to sing a word would be larger than the amplitude of the P300 evoked by an instruction to read a word aloud, and critically that the amplitude of the P300 evoked by an instruction to read a word aloud would be larger than the amplitude of the P300 evoked by an instruction to read a word silently.

Method

Participants

Twenty-seven participants (5 male, 22 female; $M_{age}=20.2$, $SD_{age}=2.2$) with no known neurological impairments, with normal or corrected-to-normal vision, and with English as a first language took part in the experiment. All of the participants were volunteers who received credit in an undergraduate course for their participation. The participants provided informed consent approved either by the Health Sciences Research Ethics Board at Dalhousie University (17 participants) or the Human Research Ethics Board at the University of Victoria (10 participants), and the study was conducted in accordance with the ethical standards prescribed in the original (1964) and subsequent revisions of the Declaration of Helsinki.

Apparatus and Stimuli

This experiment was loaded on a PC, Windows 7, running MATLAB Version 7.14 (Mathworks, Natick, USA) using the Psychophysics Toolbox Extension (Brainard, 1997). At study, a USB webcam microphone was used to automatically detect whether participants sang or

spoke on a trial prior to an auditory cue¹ and at test a standard USB game controller was used to perform a computerized memory task. All text was presented against a uniform black background in Arial size 40 font.

Four hundred and eighty words were selected from the Paivio, Yuille, and Madigan (www.math.yorku.ca/SCS/Online/paivio/) word generator. The words were all nouns, 3 to 13 letters in length with a mean of 6.92 letters and 2.30 syllables. The words had a mean Kucera-Francis (1967) word frequency of 31.84, a mean imagery rating of 4.92, a mean concreteness rating of 4.96, and a mean meaningfulness rating of 5.84. The study words were printed in red, blue, or yellow coloured font, which represented three conditions: Read Silently, Read Aloud, and Sing. For the yes/no recognition phase, all words were also printed in red, blue, or yellow coloured font, however “yes” responses to unstudied foil words (regardless of colour) were combined into a single false alarm rate.

Prior to testing each participant, the 480 words were randomly distributed into three study lists (Read Silently, Read Aloud, Sing), each consisting of 80 words, and three foil lists (i.e., items not presented at study), each consisting of 80 words.

Procedure

Prior to beginning the experiment, participants received both oral and written instructions, and were encouraged to minimize head movements and eye blinks. Participants were told that they would be presented with a study phase immediately followed by a recognition

¹ Although the USB webcam microphone automatically detected whether a participant sang or spoke before the auditory cue, an experimenter was on the other side of the wall and tracked the accuracy of the microphone by looking at the data file as it was produced on-screen and also monitoring any sounds made by the participants. If a participant coughed or cleared his/her throat and the sound triggered the microphone, the experimenter made a note and the trial was not deleted. Thus, only trials where the participant actually sang or spoke before the auditory cue were deleted.

phase. For the study phase, participants were told that they would be presented with a series of words one at a time, printed in white font that would change to a red, blue, or yellow font colour after a short delay. Each font colour represented a specific instruction condition, which was randomized across participants. For the recognition phase, participants were informed that they would be required to complete a memory test for which the instructions would appear at that time. Finally, before beginning the experiment proper, participants were told that we would need to calibrate the microphone, after which they would be required to complete both a familiarization phase and a practice phase. These two phases were designed to ensure that participants were comfortable with the three instruction conditions (Read Silently, Read Aloud, Sing) before beginning the experiment proper (see Quinlan & Taylor, 2013, for similar methodology).

Calibration. Before beginning the experiment, we calibrated the USB webcam microphone for each participant. To do this, participants were asked to say the word “banana” aloud, 10 times, while their voice was recorded by the microphone. The experimental program determined the maximum audio level of the recording; half of this value was set as the threshold for detecting singing or speaking onset during the experiment. The microphone was used in the study phase to detect whether participants sang or spoke on a trial prior to the presentation of an auditory cue.

Familiarization phase. Participants were presented with 15 trials designed to familiarize them with the three font colours (red, blue, yellow) and their associated instruction conditions (Read Silently, Read Aloud, Sing). There were five trials for each colour and its associated instruction condition, which were presented in random order. For each familiarization trial, a blank screen was presented for 500 ms followed by the verbal descriptor of the colour along with

its associated instruction condition (e.g., 'RED-Sing') at centre for 2000 ms. Both the name of the colour and the associated instruction condition were printed in the indicated colour (e.g., 'RED-Sing' was printed in red).

Practice Phase. As in the familiarization phase, participants were presented with 15 trials to ensure that participants were comfortable with the instruction conditions and also to give them practice with the three instruction conditions (Read Silently, Read Aloud, Sing). There were 15 required practice trials: five for each instruction condition, randomly intermixed. For each practice trial, a fixation cross, which was 0.6 degrees of visual angle wide, appeared for 400 – 600 ms, and was then replaced with the word “banana” written in white-coloured text, measuring 0.9 degrees of visual angle high. Here, as in the study phase, the timings of events were jittered to reduce the impact of the ERP in response to the preceding event on the event of interest (Woldorff, 1993). After 800 - 1000 ms, the word changed colour to red, blue, or yellow font to indicate the instruction for that trial (Read Silently, Read Aloud, Sing). After another 800 - 1000 ms, a 50 ms auditory tone (sine wave, 3000 Hz) cued participants to follow the instruction for that trial. For example, if the word was presented in red, the auditory cue indicated that participants should sing that word; if the word was presented in blue, the auditory cue indicated that participants should read the word aloud; if the word was presented in yellow, the auditory cue indicated that participants should read the word silently (i.e., make no overt response or mouth movement).

Participants had 2000 ms to follow the instruction before a new trial began. These practice trials were identical to those in the study phase with the exception that “banana” was the only word presented during practice. Following the presentation of the 15 practice trials, participants were given two options: (1) If they were comfortable with the instruction conditions

and their performance on the practice trials, they could proceed to the study phase, or (2) If they were not comfortable with the instruction conditions or their performance on the practice trials, they could repeat the practice phase until they reached a point where they felt comfortable to proceed to the study phase.

Study Phase. The study phase consisted of a total of 240 trials: 80 trials in each of the three instruction conditions (Read Silently, Read Aloud, Sing), presented randomly. The timing of each study phase trial was identical to those in the practice phase. On each trial of the study phase, a central fixation cross, which was 0.6 degrees of visual angle wide, appeared for 400 - 600 ms. Next, the fixation cross was replaced with a word, written in white-coloured font, which was 0.9 degrees of visual angle high. Each word was selected randomly without replacement from the Read Silently, Read Aloud, or Sing study list. After 800 - 1000 ms, the word changed colour to red, blue, or yellow font to indicate the instruction for that trial (Read Silently, Read Aloud, Sing). After 800 - 1000 ms, an auditory tone (sine wave, 3000 Hz) was presented for 50 ms to cue participants to follow the instruction for that trial. Participants had 2000 ms to follow the instruction before a new trial began. For all study trials, a USB webcam microphone was used to detect whether participants sang or spoke prior to the auditory cue. Trials for which participants sang or spoke early were marked in the EEG file and removed from subsequent analysis.

Recognition Phase. Immediately after completing the study phase, participants proceeded to the recognition phase, which consisted of a self-paced yes/no recognition test. At the beginning of the recognition phase, instructions appeared at the top of the computer screen and remained there for the duration of the phase. Using the controller, participants were instructed to press the button on the left if they recognized the word from the study trials and to

press the button on the right if they did not recognize the word from the study trials (i.e., if it was a foil word). Immediately after a button was pressed, the next trial began.

For each recognition trial, a central fixation cross (8 mm, or 0.6 degrees) appeared for 400 - 600 ms followed by a word, which was presented in red, blue, or yellow coloured font and remained on the display until a response was made. In total, there were 480 recognition trials: the 240 words presented in the study phase plus 240 foil words, randomly intermixed. The 240 items from the three study lists were printed in the same colour font as they were in the study phase: 80 study items were printed in red, 80 study items were printed in blue, and 80 study items were printed in yellow. The 240 items from the three foil lists were printed in colours that corresponded to the study phase conditions: 80 foil items were printed in red, 80 foil items were printed in blue, and 80 foil items were printed in yellow. In contrast to the study phase, in the recognition phase participants were not required to perform the instruction condition (Read Silently, Read Aloud, Sing) associated with the font colour (red, blue, yellow) of the word.

Data Collection

The experimental program recorded instruction condition (Read Silently, Read Aloud, Sing), word type (Studied, Foil), response (Yes, No), and response time. The EEG was recorded from 64 electrode locations using Brain Vision Recorder software (Version 1.20, Brain Products, GmbH, Munich, Germany). The electrodes were mounted in a fitted cap with a standard 10-20 layout and were recorded with an average reference built into the amplifier. The vertical and horizontal electrooculograms were recorded from electrodes placed above and below the right eye and on the outer canthi of the left and right eyes. Electrode impedances were kept below 20 k Ω . The EEG data were sampled at 1000 Hz, amplified (Quick Amp, Brainproducts, GmbH,

Munich, Germany), and filtered through a passband of 0.017 Hz - 67.5 Hz (90 dB octave roll off).

Data Analysis

For the behavioural data, a hit was defined as a ‘yes’ response to studied words from the Read Silently, Read Aloud, and Sing instruction conditions; a false alarm was defined as a ‘yes’ response to unstudied foil words. The yes proportions for studied words (hit rate) and unstudied foil words (false alarm rate) were computed for each participant and analyzed in a one-way ANOVA as a function of condition (Read Silently, Read Aloud, Sing, Foil).

EEG data were filtered through a (0.1 Hz – 20 Hz pass band) phase shift-free Butterworth filter and re-referenced to the average of the two mastoid channels. Next, ocular artifacts were corrected using the algorithm described by Gratton, Coles, and Donchin (1983). Subsequent to this, all trials were baseline corrected using a 200 ms epoch prior to stimulus onset. Finally, trials in which the change in voltage in any channel exceeded 10 μ V per sampling point or the change in voltage across the epoch was greater than 100 μ V were discarded. In total, less than 10% of the data were discarded due to artifacts (Silent: 8.9%; Speak: 9.2%; Sing: 8.3%). Additionally, 0.2% of the EEG segments were removed from analysis because a participant sang or spoke early (i.e., before the presentation of the auditory tone).

To examine whether the P300 amplitude would be sensitive to specific instructions (Read Silently, Read Aloud, Sing), 800 ms epochs of data (from 200 ms before stimulus onset to 600 ms after stimulus onset) were extracted from the continuous EEG for each trial, channel, and participant for all colour changes (white to either red, blue, or yellow) during the study phase. ERPs were created for each condition (Read Silently, Read Aloud, Sing) by averaging the stimulus-locked EEG data for all trials in a condition and for all participants at each channel. We

defined the P300 component for each condition as the mean deflection in the average waveforms 300 – 550 ms post stimulus at the channel where the mean deflection was maximal (Silent: C2, Read Aloud: PO4, Sing: PO4)². The time window 300 – 550 ms post stimulus was chosen based on the location of the peaks of the grand averages for each condition. P300 peaks were analyzed in a one-way ANOVA as a function of instruction condition (Read Silently, Read Aloud, and Sing).

Results

Behavioural Results

Mean “yes” response proportions were analyzed in a one-way ANOVA by condition (Sing: .617, SD .149; Read Aloud: .605, SD .161; Read Silently: .500, SD .156; Foil: .171 SD .156). There was a significant effect of condition, $F(3,78) = 151.0$, $MSe=1.17$, $p<.001$, $\eta_p^2=.936$. Planned contrasts revealed a difference between studied words and unstudied words (Sing/Read Aloud/Read Silently vs. Foil), $F(1,26) = 370.7$, $MSe = 4.378$, $p < .001$, $\eta_p^2 = .934$. Also, recall for produced words differed from that for words read silently (Sing/Read Aloud vs. Read Silently), $F(1,26) = 30.8$, $MSe = 0.329$, $p < .001$, $\eta_p^2 = .542$. Finally, and contrary to the pattern reported by Quinlan and Taylor (2013), words that were sung were no more likely to be recalled than words that were read aloud (Sing vs. Read Aloud), $F(1,26) = 0.3$, $MSe = 0.004$, $p = .6$, $\eta_p^2 = .011$. See Appendix A for an additional analysis involving the sensitivity measure, A’.

EEG Results

² We chose to quantify each component at the channel where it was maximal for each condition (Luck, 2014) – an approach that resulted in a different channel being analyzed for one of the conditions (Silent: C2, Read aloud and Sing, PO4). However, we also conducted our analyses using channel PO4 for all conditions and found the same pattern of results as reported here.

The amplitude of the P300 was calculated on a subject-by-subject basis for each instructional condition (see Figure 1 for grand average ERP waveforms). The P300 data were submitted to a three level (Read Silently, Read Aloud, Sing) one-way repeated measures ANOVA. Paralleling our behavioural data, we again observed an effect of instruction condition, $F(2, 52) = 4.97$, $MSe = 23.92$, $p = .01$, $\eta_p^2 = .83$. Planned contrasts revealed a difference between produced items (Sing: $M = 7.02 \mu V$, $SD = 3.56$; Read Aloud: $M = 6.63 \mu V$, $SD = 3.15$) and non-produced items (Silent: $M = 5.23 \mu V$, $SD = 3.01$), $F(1, 26) = 6.55$, $MSe = 68.67$, $p = .02$, $\eta_p^2 = .87$. However, there was no difference between the Sing and Read Aloud conditions, $F(1, 26) = 0.78$, $MSe = 4.11$, $p = 0.39$, $\eta_p^2 = .44$. To further probe whether there were differences between the sing and read aloud conditions, we conducted a correlation analysis on hit rate and P300 difference scores for the Sing (Sing minus Read Silently) and Read Aloud (Read Aloud minus Read Silently) conditions. The results of this analysis revealed no correlation between these variables, $r(25) = .15$, $p = .47$.

Discussion

In the present study, we provide some replication of previous behavioural findings on the production effect. Specifically, we found enhanced memory performance for produced items relative to non-produced items (Read Aloud/Sing > Read Silently). This observation is partially in line with previous work by Quinlan and Taylor (2013) who reported a production effect in memory for singing and reading aloud relative to reading silently. However, unlike Quinlan and Taylor (2013), our post hoc comparisons did not reveal any difference between memory performance for words that were read aloud and words that were sung. Mirroring our behavioural results, the current investigation revealed larger P300 amplitudes for instructions to produce words (read aloud or sing) than for instructions to read silently. As with our behavioural

data, we found no statistically reliable difference in P300 amplitude at time of encoding between the Sing and Read Aloud instruction conditions.

Previous work has linked increases in the amplitude of the P300 component with increases in distinctiveness and related increases in memory performance at test (see Fabiani & Donchin, 1995; Fabiani et al., 1990; Kamp et al., 2012; Karis et al., 1984). Our findings are thus partially, albeit not fully, consistent with the distinctiveness account of the production effect (see below). Compared to reading silently, reading aloud includes at least two distinct elements (articulation and audition) and singing may include the additional distinct elements of intensity, pitch, and/or timbre (Quinlan & Taylor, 2013; see also Forrin et al., 2012 as well as Roederer, 2008). While some neuroimaging studies have identified differential patterns of neural activation for singing and reading aloud during production (e.g., Jeffries, Fritz, & Braun, 2003; Özdemir, Norton, & Schlaug, 2006; Stewart, Walsh, Firth, & Rothwell, 2001), these findings have not been related to subsequent memory performance or to the production effect literature. Given that the amplitude of the P300 component is sensitive to distinct processing (e.g., Fabiani & Donchin, 1995; Fabiani et al., 1990; Kamp et al., 2012; Karis et al., 1984), and assuming that singing adds distinct features beyond simply reading aloud, our results do not support a distinctiveness account of the production effect. In this experiment, P300 amplitude appeared to be sensitive only to production: reading aloud and singing relative to reading silently.

The P300 component has also been identified as a measure of effort (Kok, 2001), peculiarity (Donchin & Coles, 1988), and saliency (Squires, Donchin, Herning, & McCarthy, 1977). As well, it has been shown to be sensitive to increases in cognitive load (Krigolson, Heinekey, Kent, & Handy 2012). Differences in the P300 component observed in the current study may thus reflect a range of different processes that could contribute to the production

advantage, although time-locking events to the instruction to encode, rather than to the initiation of task performance, suggests that these effects are unlikely to be associated with task effort per se (see also Forrin, Jonker, & MacLeod, 2014).

The Distinctiveness Account of the Production Effect, Memory, and the P300

Surprisingly, we were unable to completely replicate the results of Quinlan and Taylor (2013). In particular, while we observed a production effect – both read aloud and sung words were more likely to be recognized than silently read words – we observed no difference between the recall of sung and read aloud words. Further, we found no difference in P300 amplitude between the sing and read aloud conditions. Thus, on the surface, our results do not appear to support the distinctiveness account of the production effect. According to Quinlan and Taylor's (2013) interpretation, the distinctiveness account of the production effect would predict that singing includes more distinct items than reading aloud, thereby providing the basis for the memorial advantage previously observed for sung items. Why was this advantage not observed in the present experiment? The simplest answer is that it may be that some (or all) of our participants did not sing the words in a manner that included the distinct elements of intensity, pitch, and/or timbre. Indeed, the production effect has been associated with greater engagement of the motor system (e.g., Bodner & Taikh, 2012). If in the present experiment the motor system was similarly engaged between the Sing and Read Aloud conditions, then one would not predict a difference between these conditions.

Another explanation for our failure to fully replicate Quinlan and Taylor (2013) centers on a methodological difference between the two experiments. Here, for ERP methodological reasons, we separated the instruction to sing, read aloud, or read silently from the instruction to perform the action. We chose to do this to remove possible motor contamination in the ERPs

from the instruction cue. However, by doing so, we created an instruction cue that provided intention but not an actual command to perform. Indeed, it could be that the temporal separation between intention and production resulted in a reduced distinctiveness for sung words. In other words, the distinctiveness generated by the instruction is a separate cue to memory than the distinctiveness generated by the processing engaged to execute that instruction (which becomes less distinct the longer the delay to enactment). Both are important in the effect observed by Quinlan and Taylor (2013), but the latter became weaker in the present study, given the necessities of design.

Although our findings offer a potential neural marker for studying the production effect, our study is not without limitations. For example, it is possible that the effect of instruction condition on the P300 component is separate from and unrelated to the effect of instruction condition on memory performance. However, given that numerous studies (e.g., Fabiani & Donchin, 1995; Fabiani et al., 1990; Kamp et al., 2012; Karis et al., 1984) have reported a direct relation between the amplitude of the P300 component at encoding and memory performance at test, we believe that this is also likely true in our study; future work should explicitly explore this possibility. Participants were told that all encoding trials would be monitored using a microphone and that if they spoke or sang early the current trial would be discarded. While this warning was highly effective (very few trials were discarded due to early singing/speaking) we may have inadvertently impacted the encoding of sung/spoken words in some way.

In conclusion, we observed better memory recall and increased P300 amplitudes for encoding instructions to read aloud and sing relative to an encoding instruction to read silently. Although our results reflect a production effect – produced items had a memorial advantage and increased P300 amplitudes relative to non-produced items – our results are not fully in line with

the distinctiveness account of the production effect (at least if it is assumed that, relative to speaking, singing increases the number of distinctive features). Rather, our results highlight that the P300 ERP component, at least in the context of the present experiment, is sensitive to production. Further research at the neural level should put the role of distinctiveness – and indeed of other potential underlying mechanisms – in the production effect to stringent test.

Acknowledgements

Support for this study was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) through a Canadian Graduate Scholarship to CDH, a Vanier Scholarship to CKQ, funds from the Neuroeducation Network at the University of Victoria, and a Discovery Grant to TLT and OEK.

References

- Bodner, G. E., & Taikh, A. (2012). Reassessing the basis of the production effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1711-1719.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433-436.
- Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, *110*, 24-30.
- Castel, A. D., Rhodes, M. G., & Friedman, M. C. (2013). Predicting memory benefits in the production effect: The use and misuse of self-generated distinctive cues when making judgments of learning. *Memory & Cognition*, *41*, 28-35.
- Crites, S. L., Devine, J. V., Lozano, D. L., & Moreno, S. (1998). Event-related potentials and serial position effects in a visual probe recognition task. *Psychophysiology*, *35*, 293-304.
- Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*, 155-161.
- Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory and Cognition*, *24*, 523-533.
- Donchin E. Surprise!....Surprise? *Psychophysiology* 1981;18:493–513.
- Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioural and Brain Sciences*, *11*, 357–374.
- Donchin, E., Karis, D., Bashore, T.R., Coles, M.G.H., & Gratton, G. (1986). Cognitive psychophysiology: systems, processes, and applications. In: M.G.H. Coles, E. Donchin, and S. Porges (Eds). *Psychophysiology: Systems, processes, and applications* (pp. 244-267). New York: The Guilford Press.

- Fabiani, M., Karis, D., & Donchin, E. (1990). Effects of mnemonic strategy manipulation in a Von Restorff paradigm. *Electroencephalography and Clinical Neurophysiology*, *75*, 2-35.
- Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 224-240.
- Fawcett, J. M. (2013). The production effect benefits performance in between-subjects designs: A meta-analysis. *Acta Psychologica*, *142*, 1-5.
- Fawcett, J. M., Quinlan, C. K., & Taylor T. L. (2012). Interplay of the production and picture superiority effect: A signal detection analysis. *Memory*, *20*, 655-666.
- Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, *40*, 1046-1055.
- Forrin, N. D., Jonker, T. R., & MacLeod, C. M. (2014). Production improves memory equivalently following elaborative vs. non-elaborative processing. *Memory*, *22*, 470-480.
- Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468-484.
- Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *11*, 534-537.
- Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 3-25). New York: Oxford University Press.
- Icht, M., Mama, Y., & Algom, D. (2014). The production effect in memory: Multiple species of distinctiveness. *Frontiers in Psychology*, *5*.
- Jeffries, K. J., Fritz, J. B., & Braun, A. R. (2003). Word in melody: An H¹⁵₂O PET study of brain

- activation during singing and speaking. *NeuroReport*, *14*, 749-754.
- Kamp, S. M., Forester, G. R., Murphy, A. R., Brumback, T., & Donchin, E. (2012). Testing a distinctiveness explanation of the primacy effect in free recall using event-related potentials. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 539-544). Austin, TX: Cognitive Science Society.
- Karis, D., Fabiani, M., & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology*, *16*, 117-216.
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557-577.
- Krigolson, O. E., Heinekey, H., Kent, C. M., & Handy, T. C. (2012). Cognitive load impacts error evaluation within medial-frontal cortex. *Brain Research*, *1430*, 62-67.
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Luck, S. (2014). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 671-685.
- Özdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *NeuroImage*, *33*, 628-635.

- Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1543-1547.
- Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology*, *67*, 360-371.
- Patterson, J. V., Pratt, H., & Starr, A. (1991). Event-related potential correlates of the serial position effect in short-term memory. *Electroencephalography and Clinical Neurophysiology*, *78*, 424-437.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128-2148.
- Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory*, *21*, 904-915.
- Richler, J. J., Palmeri, T. J., & Gauthier, I. (2013). How does using object names influence visual recognition memory? *Journal of Memory and Language*, *68*, 10-25.
- Roederer, J. G. (2008). *The physics and psychophysics of music: An introduction* (pp. 1-242). New York: Springer.
- Rushby, J. A., Barry, R. J., & Johnstone, S. J. (2002). Event-related potential correlates of serial-position effects during an elaborative memory test. *International Journal of Psychophysiology*, *46*, 13-27.
- Squires, K. C., Donchin, E., Herning, R. I., & McCarthy, G. (1977). On the influence of task relevance and stimulus probability on event-related-potential components. *Electroencephalography and Clinical Neurophysiology*, *42*, 1-14.
- Stewart, L., Walsh, V., Frith, U., & Rothwell, J. C. (2001). Transcranial magnetic stimulation

produces speech arrest but not song arrest. *Annals of the New York Academy of Sciences*, 930, 433-435.

Woldorff, M. G. (1993). Distortion of ERP averages due to overlap from temporally adjacent ERPs: Analysis and correction. *Psychophysiology*, 30, 98-119.

Appendix A

Additional Analysis of Word Recall

Because foil words were presented in the study instruction colours, we were able to compute three separate foil false alarm rates. Consistent with Quinlan and Taylor (2013), these independent false alarm rates were used to calculate sensitivity (A' : Donaldson, 1996) for the study lists presented in the corresponding colour (e.g., if red signaled a Sing instruction at study, the foil items presented in red provided the false alarm rate for calculations made in the Sing condition: see Quinlan & Taylor, 2013). Lower A' values represent lower sensitivity and higher A' values represent greater sensitivity (a value of .50 represents chance performance). A' scores were analyzed in a three level (Read Silently, Read Aloud, Sing) one-way repeated measures ANOVA. This analysis partially replicated the findings of Quinlan and Taylor (2013) in demonstrating an effect for instruction condition, $F(2, 52)=19.4$, $MSe=.030$, $p<.001$, $\eta_p^2=.4273$. Planned contrasts revealed a difference between the Sing ($M=.82$, $SD=.06$) and Silent ($M=.76$, $SD=.06$) conditions, $t(26) = 4.59$, $p < .001$, and between the Read Aloud ($M=.82$, $SD=.06$) and Silent conditions, $t(26) = 5.25$, $p < .001$. However, unlike Quinlan and Taylor (2013), we observed no difference between the Sing and Read Aloud conditions, $t(26) = 0.09$, $p = .93$.

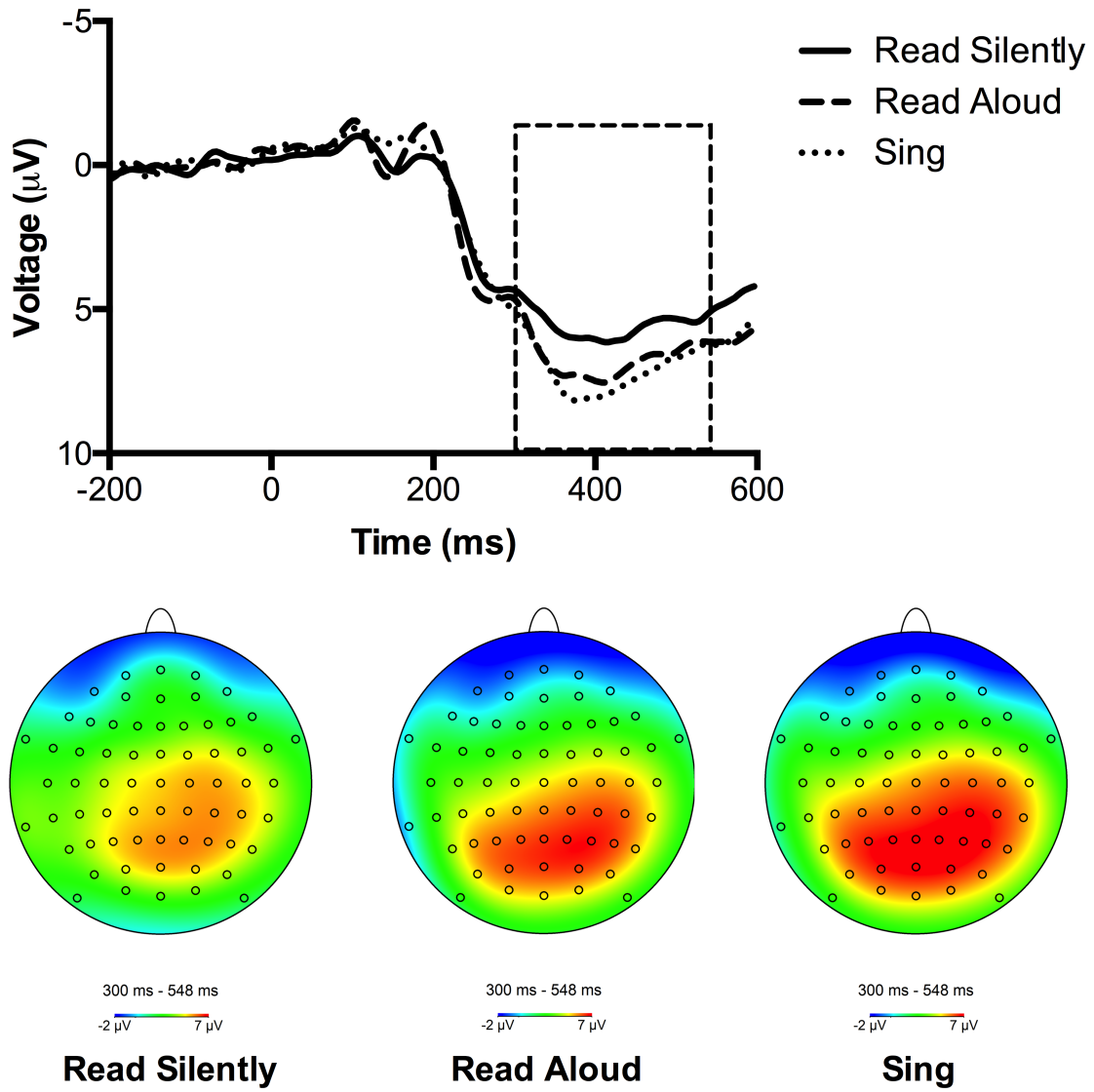


Figure 1. P300 in response to instruction condition (Read Silently, Read Aloud, Sing) with scalp topographies. Dashed line shows the time range for analysis (300 ms – 550 ms post instruction).