**From Thought to Senses: Assessing the Presence of a Relationship Between the Generation Effect and Multisensory Facilitation**

Michaela Ritchie and Jonathan Wilbiks

Department of Psychology, University of New Brunswick Saint John

**Author Note**

Department of Psychology, University of New Brunswick Saint John

The present study was reviewed by the Research Ethics Board (REB File #012-2024) at

the University of New Brunswick.

Correspondence concerning this article should be addressed to Michaela Ritchie,

University of New Brunswick Saint John, 100 Tucker Park Rd, Saint John, NB E2K 5E2,

Canada.

Email: [michaela.ritchie@unb.ca](mailto:michaela.ritchie@unb.ca)

**Abstract**

The proposed study will investigate the relationship between the generation effect, a memory advantage for self-generated verbal information, and the multisensory facilitation effect, a phenomenon wherein congruent sensory inputs enhance cognitive processing. Although extensive research has explored the generation effect, the multisensory aspect of this task has been overlooked. The proposed study aims to determine if the generation effect is modulated by the multisensory facilitation effect. Multisensory facilitation involves the brain's ability to efficiently process information from diverse sensory sources simultaneously, enhancing cognitive performance. The proposed study will employ a 2 (Task Type: generate, read) X 3 (Sensory Modality: auditory, visual, audiovisual) factorial design, utilizing word pair lists and recognition tests. It is hypothesized that the multisensory facilitation effect will amplify the generation effect, particularly in the audiovisual condition. This research contributes to understanding the interplay between self-generation, multisensory processing, and memory, providing insights for cognitive mechanisms, and having implications for educational settings. The proposed study will address a gap in existing literature by examining the multisensory component of generation tasks and extends the current understanding of multisensory learning by incorporating verbal stimuli as the visual component.

*Keywords:* generation effect, multisensory facilitation effect, audiovisual learning, verbal memory, recognition

**From Thought to Senses: Assessing the Presence of a Relationship Between the Generation Effect and Multisensory Facilitation**

The generation effect refers to the advantage in memory performance for verbal information when it is generated by one’s own mind, rather than passively read (Slamecka & Graf, 1978). In previous studies, an experimenter would cue participants by presenting them with a rule and asking them to generate a target word. The ability to recall a greater number of generated target words than words that are read, constitutes the generation effect (Slamecka & Graf, 1978). The magnitude of the effect is calculated by subtracting the number of correctly remembered read words from the number of correctly remembered generated words. The phenomenon is a robust effect, improving both item memory (word-specific) and context memory (information surrounding the word), compared to reading (McCurdy et al., 2020). While a substantial body of research has focused on the widely recognized phenomenon known as the *generation effect*, which demonstrates that self-generated verbal information is better retained for later recall or recognition than passively read information (e.g., Bertsch et al., 2007; McCurdy et al., 2020; Slamecka & Graf, 1978), the multisensory aspect of this task has thus far been neglected in research. Although the generation task inherently involves a multisensory process of interpreting visual stimuli, and vocalizing the text to produce its auditory representation, this particular aspect of the task has not been explicitly investigated in research. One specific aim of the proposed study is to examine whether the generation effect is modulated by the multisensory facilitation effect.

**Multisensory Facilitation Effect**

Contemporary research in cognitive psychology, educational psychology, and neuroscience, has shed light on the adaptability of the human brain to efficiently process multisensory information, a capacity finely tuned to accommodate humans’ multisensory way of living (Shams & Seitz, 2008). This intricate process encompasses various stages of sensory perception, from the brain’s initial reception of information to the processing within early sensory cortical areas and the formation of associations among other distinct cortical regions (Alais et al., 2010; Falchier et al., 2002; Rockland & Ojima, 2003). Although the processing of information from the senses was initially thought as a unisensory process (e.g., Penfield & Rasmussen, 1950; Treisman & Gelade, 1980), the presence of multisensory activations and complex neural networks throughout the brain have since been discovered (Burton et al., 2002; Falchier et al., 2002; Ghazanfar & Schroeder, 2006; Rockland & Ojima, 2003; Röder et al., 2002). This intricate connectivity underlines the efficiency with which the brain integrates and processes information from diverse sensory sources simultaneously, which allows humans to respond and adapt to the multiple sources of information in the environment and reduce cognitive load by pairing *congruent stimuli* together and processing them using overall fewer cognitive resources than when processing incongruent stimuli (Fahey et al., 2018; Shams & Seitz, 2008). Shams and Seitz (2008) define congruent stimuli as “[…] the relationship between stimuli that are consistent with the prior experience of the individual or relationships between the senses found in nature” (p.2). By this definition, arbitrary associations between auditory and visual stimuli will not bring about a multisensory facilitation effect. Moran et al. (2013) elaborated on this proposal: if two stimuli are paired for the purpose of a specific task that are not semantically congruent in nature (e.g., an auditory input of a dog barking paired with an image of a lion), it will still be possible to integrate these two stimuli during encoding, but will likely result in only partial or occasional integration at the subconscious level later on, leading to deleterious effects of multisensory input of stimuli in such a scenario (Moran et al., 2013; Shams et al., 2011). Congruent pairings of stimuli are developed within an individual’s semantic network through repeated exposure or experience to the pairings in nature. When such ecologically associated visual and auditory stimuli are presented together, processing of either stimulus will be facilitated as a result of an enhanced, and more elaborate representation of the information across the semantic network and multiple cortical regions in the brain (Moran et al., 2013; Shams et al., 2011). The integrated perception of percepts from multiple senses (i.e., multisensory processing of a single experience) is contingent on *temporal binding*, whereby information acquired from each sense be provided within a certain window of time so that the percepts may be unified as a coherent perception of the stimuli (Meredith et al., 1992; Wallace et al., 1996; Wallace & Stevenson, 2014). Failure to adhere to the temporal binding window could result in the perception of two unintegrated unisensory experiences in lieu of one multisensory percept (Wallace et al., 2004; Stevenson et al., 2012). The concept of the temporal binding window explains perceptual illusions like the ventriloquism effect. In this phenomenon, the observer incorrectly attributes auditory stimuli to the ventriloquist's dummy due to synchronization between the dummy’s lip movement and the accompanying speech sounds (Howard & Templeton, 1966). This underscores that when auditory and visual cues coincide within a specific temporal window, an illusion is created, leading to the perception of a unified and cohesive experience (Wallace et al., 2004). The precise time period during which stimuli from various modalities are prone to perceptual binding is debated as being as little as 50–100 ms (Slutsky & Recanzone, 2001), 200 ms (Lewald et al., 2001) to as great as 800 ms (Wallace et al., 2004). Ultimately, the simultaneous presentation of multiple stimuli ensures that auditory and visual stimuli are perceived concordantly, as the temporal window between stimuli provide statistical cues to their relationship (Wallace & Stevenson, 2014).

A pivotal aspect of multisensory processing is its apparent facilitation effect, wherein unisensory items exhibit enhanced recognition when paired with congruent sensory inputs. Seitz et al. (2006) demonstrated that participants who learned to recognize visual stimuli with congruent auditory stimuli required approximately 60% fewer training sessions than those who studied visual stimuli in isolation. In early discussion of the outcomes for multisensory processing, it was debated that the pairing of an auditory stimulus with a visual stimulus provided an alerting effect, whereby the visual stimulus became more prominent in the mind as a result of increased attention from an auditory alerting stimulus that was paired with it during encoding of the visual stimulus. If the introduction of an auditory stimulus improves the processing of visual stimuli due merely to an alerting effect, then this effect should be noticeable for both congruent and incongruent stimuli. Seitz et al. (2006) contested the proposal of a mere alerting effect by training individuals to respond to visual stimuli in a unisensory processing condition, as compared to a multisensory condition where the auditory stimuli paired was either congruent or incongruent. A facilitation effect of multisensory processing was only observed for the congruent multisensory condition, supporting the notion that when semantically congruent information is presented from multiple sources, the representation of a unisensory piece of information becomes richer as a result of being partially represented in other connected areas in the brain (Friston, 2005; Seitz et al., 2006). These connections are established through the multisensory learning conditions (Seitz et al., 2006) and any aspect of the multisensory representation can be elicited by merely activating a unisensory component (Friston, 2005; Rao & Ballard, 1999). Thus, the representation of unisensory components is also improved as a result of an enhancement in its representation, which is distributed across a multisensory network (Shams & Seitz, 2008).

Multiple studies have demonstrated that multisensory processing is related to improved accuracy, precision, and reaction times in various tasks (Alais & Burr, 2004; Ernst & Banks, 2002; Gingras et al., 2009). Such findings support a multisensory processing facilitation effect, whereby performance of cognitive tasks for multisensory information, or for unisensory components of information learned in a multisensory context, are enhanced (Kim et al., 2008). The multisensory facilitation effect highlights the brain's capacity to create more elaborate representations of information across different regions when multiple senses are engaged (Kim et al., 2008).

Although much of the research on multisensory facilitation is centered around non-verbal stimuli, as has been reviewed in this section, a more recent focus on the verbal stimuli has also come about in the literature. Kim (2021) found that the pairing of e-book reading with the same passage in auditory format (audiovisual), improved listening comprehension compared to the e-book (visual) and audiobook (auditory) conditions alone. Interestingly, reading comprehension was only better in the audiovisual condition compared to the auditory condition only, but not than the visual only condition. In fact, the auditory condition yielded worse reading comprehension than the visual condition as well (Kim, 2021). A more detailed discussion of facilitation of multisensory verbal information processing is presented in the following section.

**Congruency as a Determinant of Facilitation**

The concept of congruency between sensory stimuli plays a central role in triggering the multisensory facilitation effect. Shams and Seitz (2008) defined congruent stimuli as those that align with an individual's prior sensory experiences or with the natural relationships between sensory modalities. Moran et al. (2013) further elaborated on the critical role of congruency, highlighting that incongruent pairings may only result in partial or occasional integration at the subconscious level, potentially leading to deleterious effects on multisensory information processing.

The proposed mechanism behind multisensory learning involves neural modulation, where signals originating from neurons of one modality influence the activation of neurons in another modality (Barutchu et al., 2013; Seitz & Dinse, 2007). Through a pairing process of visual and auditory stimuli, the representation of unisensory information becomes interleaved with the representation of the multisensory information, which in turn will activate unisensory information from the other sense (Barutchu et al., 2013; Seitz & Dinse, 2007). These connections are established through the multisensory learning conditions (Seitz et al., 2006) and any aspect of the multisensory representation can be elicited by merely activating a unisensory component (Friston, 2005; Rao & Ballard, 1999). Thus, the representation of unisensory components is also improved as a result of an enhancement in its representation, which is distributed across a multisensory network (Shams & Seitz, 2008). According to Seitz and Dinse (2007), perceptual learning requires the activation of neurons to surpass a certain threshold. This theory suggests that modulating the activity in one sensory modality, such as sound influencing visual activation or vice versa, can facilitate learning and information processing across modalities.

The discussion surrounding the congruency effect becomes more nuanced in consideration of the multisensory input of verbal stimuli. Kalyuga (2000) proposed that the negative impact on learning arises from the substantial cognitive load incurred when concurrently processing multiple related sensory sources of information. Consequently, the utilization of redundant sensory modalities may prove ineffective for learning in some cases (Kalyuga, 2000). Other researchers have found that there is no effect of multimodal learning; for example, Rogowsky et al. (2016) observed no differences in comprehension for individuals in either unisensory condition (visual, auditory) compared to the multisensory condition (audiovisual).

Diao and Sweller (2007) assessed the impact of simultaneous verbatim spoken presentation alongside written materials on learning to read in English for non-native speakers. Two instructional techniques were compared: reading materials presented in written form only and reading materials presented in written form, accompanied by auditory articulation of the materials. In their study, concurrent spoken presentation hindered reading comprehension compared to the written-only condition. The group exposed to the audiovisual learning condition exhibited lower lexical knowledge than the group in the visual reading condition. Moreover, at the text comprehension level, the former group reported a higher cognitive load and inferior understanding and recall of the main ideas of the text (Diao & Sweller, 2007).

The *cognitive load theory* posits that learning involves accumulating domain-specific knowledge in long-term memory (Plass et al., 2010). The *borrowing principle* suggests that learners benefit from accessible knowledge, and if such knowledge is lacking, the *randomness as genesis* principle advocates for random generation and testing (Diao & Sweller, 2007; Plass et al., 2010). To avoid the need for novices to randomly infer connections between written and spoken text, information should be presented in a way that reduces this demand on working memory (Diao & Sweller, 2007). Simultaneously presenting written and spoken text to novices without relevant knowledge may lead to random testing of hypotheses, with little gain in learning to read, and increased demands on working memory.

Other findings have since contradicted this proposal and thus indicate that multimodal learning improves comprehension among both first language learners (e.g., Kim, 2021) and second language learners (e.g., Mohamed, 2018). For example, Mohamed (2018) found that listening comprehension was improved among individuals in an audiovisual learning condition, as compared to those in either unimodal condition, when learning English as a foreign language. Many other studies have reported an advantage in multimodal learning when compared to unisensory learning. As aforementioned, Kim (2021) observed an advantage for audiovisual learning on listening comprehension compared to visual or auditory learning, and an advantage in reading comprehension compared to the auditory learning condition. Further, Liu et al. (2019) found that the audiovisual learning of verbal stimuli was related to improved reading comprehension compared to either auditory or visual learning modalities alone.

Chang (2009) assessed second language learning using either unimodal or multimodal instructional techniques. Although the results indicated a modest 10% improvement with the audiovisual mode, students exhibited a marked preference for this mode. Most students perceived that listening in the audiovisual mode facilitated easier comprehension, shortened the perceived duration of tasks, enhanced the appeal of learning, and improved their overall sustained attention (Chang, 2009). A notable takeaway from this study, as compared to the disadvantages reported by Diao and Sweller (2007), is the appeal of audiovisual learning among learners, which may encourage long-term learning (Chang, 2009). Kartal and Simsek (2017) reported similar results, and their qualitative findings indicated that visual learning of verbal stimuli accompanied by auditory articulation of that stimuli motivated students to continue learning.

Overall, the literature regarding the advantage of multimodal learning for verbal stimuli is mixed, with many asserting its benefits (Chang, 2009; Kartal & Simsek, 2017; Liu et al., 2019; Mohamed, 2018), and others finding a null (Rogowsky et al., 2016) or negative effect (Diao & Sweller, 2007; Kalyuga, 2000). Given the pervasiveness of multimodal learning in academic settings, these indeterminate results underline the need for further examination.

**Discrediting Mere Alerting Effects**

Early discussions of multisensory processing debated the possibility of an alerting effect, suggesting that pairing an auditory stimulus with a visual stimulus could enhance visual processing, and that this enhancement was a result of augmented attention toward the visual stimuli, alerted by the auditory component. If this explanation were true, then the congruency of the auditory stimulus to the visual stimulus should be irrelevant; that is, an alerting effect should persist even when the auditory stimulus is semantically unrelated to the visual stimulus. Seitz et al. (2006) conducted experiments to test this hypothesis, and their results revealed that the facilitation effect was specific to congruent multisensory conditions. Moran et al. (2013) contended that when stimuli are incongruent, the brain seeks to infer independent causes, leading to independent representations of its unisensory components in their respective cortical regions, with no linkage between components being formed. In their study, recognition of images was facilitated only by congruent pairing of auditory stimuli during encoding. At the recognition stage of encoding, the activation of a unisensory component will initiate the process of activating an entire cortically-distributed network of information across multiple senses (Moran et al., 2013). These findings challenge the notion of a mere alerting effect and suggests that semantically congruent information presented from multiple sensory sources enriches the representation of unisensory information through interconnected brain regions. Multiple neuroimaging studies have corroborated this explanation: cortical regions corresponding to both the sensory modality of a previously learned unisensory recognition item and the modality of the originally paired multisensory set were activated upon recall (Gottfried et al., 2004; Nyberg et al., 2000; Wheeler et al., 2000). Once a connection between two components (specifically auditory and visual components) has been established, the introduction of partial data triggers the anticipation of the absent component(s) that were encoded alongside the primary stimulus (Friston, 2005; Moran et al., 2013). Such an explanation is reminiscent of a type of distributed object recognition process in cognitive psychology, redintegration(Haxby et al., 2001; Nyberg et al., 2000; Rolls et al., 1997; Shams & Seitz, 2008). *Redintegration* is the mechanism through which the multisensory encoding of memories may rely on a distributed network of perceptual and memory traces. In this process, the activation of a single node is adequate to elicit the activation of the complete representation of an object (Haxby et al., 2001; Rolls et al., 1997). A similar mechanism appears to underpin the outcome by which unisensory components are better learned and remembered as a result of multisensory representations (Moran et al., 2013; Sheffert & Olson, 2004; Von Kriegstein & Guiraud, 2006), as is suggested by the reviewed literature in this section.

**The Role of Semantic Relatedness of Stimuli**

Memory is not confined to specific localized regions but rather involves a widespread network of interconnected neurons, producing representations across a distributed network (Fuster, 1997; Gardiner & Hampton, 1985; Kinoshita, 1989). The representation of information is distributed across multiple semantic connections, establishing and reinforcing neural connections across distinct cortical regions (Fuster, 1997; McNamara, 2005; Vandenberghe et al., 1996). With regard to the recall of information, semantically related pieces of information benefit from a more robust and interconnected representation within this distributed network, leading to enhanced recall compared to the recall of unrelated information, which is a more effortful process that involves the activation of multiple individual regions (Heisz et al., 2014; Morais et al., 2013). Similar to the mechanisms underlying the multisensory facilitation effect, the activation of one element contributes to the activation of related elements, enhancing the efficiency of information retrieval. For this reason, an additive effect of multisensory facilitation and semantic relatedness should contribute to superior recall of related information that is processed across multiple sensory modalities (Heisz et al., 2014; Morais et al., 2013).

Semantically related information shares commonalities in meaning, creating a network of interconnected concepts in the brain (Gardiner & Hampton, 1985; Kinoshita, 1989; Slamecka & Graf, 1978). The distributed network model suggests that the neural representation of a concept is not isolated but involves the activation of a network of nodes that are related to that concept, and the activation of one element spreads across multiple related nodes of information simultaneously to process and retrieve related information more efficiently and successfully (Fuster, 1997; Heisz et al., 2014; Morais et al., 2013). This interconnected activation creates a broader and more intricate pattern of neural activity, which contributes to a richer and more robust representation of the information (Vandenberghe et al., 1996). As a result, more elaborate retrieval pathways are established, and the distributed nature of memory allows for multiple pathways to access the same information (Morais et al., 2013).

Semantically related information benefits from having more efficient retrieval pathways due to the shared connections between associated concepts. This facilitates quicker and more accurate recall (Heisz et al., 2014; Morais et al., 2013). The semantic activation model suggests that the brain's ability to recall information is optimized when that information is semantically related, which has been found to underlie the magnitude of the generation effect (Kinoshita, 1989; McCurdy et al., 2020).

**Congruency vs. Redundancy**

The literature surrounding benefits of congruency span the literature assessing both verbal (Fairhall & Macalusco, 2009; McGurk & MacDonald, 1976; Summerfield & McGrath, 1984; Wyk et al., 2010) and non-verbal (Moran et al., 2013; Shams et al., 2011) stimuli. In a multisensory context of learning, it is possible to have a primary medium of information transmission (e.g., visual), which is complemented by different, but semantically consistent information through another sensory modality (e.g., auditory information). Although congruency refers to the complementary nature of stimuli from multiple sensory sources (Kim, 2021), it is important to distinguish its beneficial effects from the effects of redundancy. In a multisensory context of verbal stimuli, redundancy refers to the same—not similar—verbal information being presented through both the visual and auditory sensory modalities (Kalyuga, 2000; Mayer, 2001). When the same information is being processed through two different sensory modalities, processing demands may be increased, increasing cognitive load.

An example of a congruent input would be the input of a cue word through the visual medium, accompanied by its auditory articulation at the same time (redundancy), and the output of the generated target word is auditorily produced while writing it down (congruency).

The incorporation of multiple sensory modalities demonstrates benefits in navigating the environment, as it serves to mitigate sensory uncertainty through the assimilation of diverse sensory inputs (Frassinetti et al., 2002; McDonald et al., 2000).

**Hypotheses**

The following hypotheses were derived from a comprehensive examination of relevant literature.

Hypothesis one (H1) posits the presence of a multisensory facilitation effect, wherein the maintenance of the multisensory component in generation tasks is anticipated to enhance subsequent recognition of target items (Fahey et al., 2018; Moran et al., 2013; Shams & Seitz, 2008). For unisensory generation tasks, the resulting memory advantage will be reduced, compared to the multisensory generation tasks. H1 proposes that the magnitude of the generation effect will increase when more than one sensory modality is engaged during encoding, specifically, during the audiovisual condition, compared to either unisensory reading task (auditory or visual). This hypothesis was formed in consideration of research that indicated an objective advantage and subjective preference for learning through a multisensory condition as opposed to a unisensory learning condition (e.g., Chang, 2009; Kartald & Simsek, 2017; Kim, 2021; Liu et al., 2019; Mohamed, 2018).

H1a: Multisensory engagement will enhance recognition of target items during generation tasks.

H1b: Audiovisual presentation will yield higher recognition of target items than auditory or visual conditions, specifically in generation tasks.

Hypothesis two (H2) outlines the expectation that, overall, recognition outcomes, as reflected by the proportion of correctly recognized items and in confidence about those decisions, will demonstrate a substantial advantage for generation tasks as opposed to the reading tasks (Bertsch et al., 2007; McCurdy et al., 2020; Slamecka & Graf, 1978). The discrepancy in learning outcomes is expected to be more pronounced for the audiovisual generation tasks compared to the unisensory reading tasks, as proposed by H1 (McCurdy et al., 2020; Moran et al., 2013; Shams & Seitz, 2008).

H2a: Generation tasks will yield higher recognition outcomes and confidence ratings compared to reading tasks.

H2b: Audiovisual generation tasks will exhibit a greater discrepancy in outcomes compared to unisensory reading tasks.

**Method**

**Participants**

The proposed study has been approved by the University of New Brunswick’s institutional review board (REB # 2024-012). Using G\*Power (Faul et al., 2009), and specifying “Repeated measures, between factors, within factors” per our analysis plan proposed below, it was determined that 69 participants are required to achieve power of 0.90 to detect a medium effect size, which would be consistent with past work on the phenomenon (see Bertsch et al., 2007, for a review). To ensure that the order of the tasks is counterbalanced evenly across participants, a target sample size of 72 participants was determined to be suitable for the proposed study. To ascertain that the appropriate sample size is achieved for analysis, participants will be excluded from analysis if they terminate the study early, there is a failure of any equipment or materials during testing, or any of the tasks are incorrectly completed. Such inclusion criteria will be checked by the researcher during the testing phase of the study. Participants will only be included in the analysis if all tasks are appropriately completed as outlined in the procedure.

**Material**

***Word Pair Lists***

For the proposed study, participants will use a word pair list containing 120 word pairs (synonyms). For each word pair, the target word will be a synonym for the cue word (e.g., chilly-cold). Using a counterbalanced design, some participants will first be exposed to the completed word pair lists under the reading conditions, and the other participants will first receive the incomplete word pair lists—which will contain only the cue words and the first letter of each target word—for the generation tasks.

***Encoding Conditions***

For all tasks, participants will be exposed to word pairs one at a time at a 4 s pace (see Appendix). Each word pair (complete or incomplete) will be presented on a 3 x 5” index card. For incomplete word pairs, an enumerated piece of paper will be provided so that participants may record their generated responses in the visual generation condition. For all tasks, participants will be facing in the opposite direction of the researcher, to eliminate potential extralingual influences during the study, such as facial expressions, gestures, body language, or eye contact, which may influence interpretation (Fairhall & Macaluso, 2009).

**Read.** Participants will be given complete word pairs, one at a time, presented each on 3 x 5” index cards. In the visual condition, participants will complete this task by reading each word pair silently without auditorily vocalizing the words. In the auditory condition, both words of the word pair will be read aloud by the researcher at a 4 s pace per pair. In the audiovisual condition, participants will be instructed to read both words aloud, only once, before moving onto the next word pair.

**Generate.** On each 3 x 5” index card, the cue word will be presented, accompanied by the first letter of the target word. In the visual condition, participants will read the cue word silently, and record the target word on an enumerated sheet of paper. In the auditory condition, the first word of the word pair will be read aloud by the researcher, followed by the first letter of the target word, at which point, the participant shall utter the target word aloud. The target words uttered by the participant will be recorded by the researcher.

**Procedure**

The proposed study will follow a 2 (Task type: read, generate) X 3 (Sensory condition: auditory, visual, audiovisual) factorial design, where task type will be assigned within subjects, and sensory conditions will be assigned between subjects. The order of tasks (read, generate) will be counterbalanced across participants. Further, of the total 120 word pairs, 60 word pairs will be assigned to the reading condition and 60 word pairs will be assigned to the generation condition, in a counterbalanced order across participants, so that any observed difference can be confidently attributed to the manipulation of task type.

***Encoding Phase***

Each word pair will be presented at a 4 s pace, which will be assured using an interval timer which signals a tone every 4 s to indicate to the participant to move to the next word pair.

***Distraction Phase***

Following each encoding phase, each participant will be instructed to continually subtract seven from a given number on a blank sheet of paper, for 30 seconds.

***Visual Recognition Condition***

Participants will be given a written recognition test on the target words. If the participant did not correctly generate the anticipated target word, their score will be adjusted to reflect the words that were actually encoded. For each question, three words of the same letter length—two lures and one target word—will be presented. The participant must circle which of the three options they recognize and they will also be asked to rate, on a scale from 1-10, how confident they are about their choice (1 = *not confident at all*, 5 = *extremely confident*). There will be a total of 120 questions on the recognition test, reflecting the total number of possible target items to be encoded.

**Proposed Analysis**

To assess whether an aspect of the generation effect can be accounted for by the multisensory facilitation effect of the task, the following research design is proposed: A 2 (Task Type: generate, read) X 3 (Sensory Modality: auditory, visual, audiovisual), factorial analysis of variance (ANOVA) will be conducted, with task type assessed between subjects, and sensory modality within subjects. To test whether H2 is also supported by confidence ratings, confidence ratings will be assessed using an independent samples *t*-test, with the grouping variable being task type (read, generate) and the outcome variable being confidence. The researchers will utilize Bayes' analysis as a follow-up to non-significant ANOVA findings to evaluate the strength of evidence for the null hypothesis and quantify the likelihood of the observed data given both the null and alternative hypotheses. This approach will offer a more nuanced interpretation of the data and provide insights into the presence or absence of effects.

**Implications**

Despite having a vast body of literature to rely on when examining the effect of self-generation on memory under a host of different conditions (for reviews of this literature, see Bertsch et al., 2007; McCurdy et al., 2020), the multisensory component inherent to most generation tasks has been neglected as an area of focus in this research. To help understand the mechanisms underlying the generation effect, assessing the multisensory component of the task will be necessary to elucidate whether generation tasks are exemplary of the multisensory facilitation effect, or whether the generation effect is as substantial when the multisensory component of generation tasks is removed. The importance of clarifying the multisensory role of encoding under such conditions extends beyond the purpose of understanding the mechanisms which underlie the generation effect. Beyond this, the findings of this study will contribute to the growing literature on multisensory learning, and will contribute evidence toward the debate regarding when a multisensory facilitation effect may be observed. Further, existing research supporting the multisensory facilitation effect examines the relationship between auditory and visual processing using auditory stimuli and images as visual stimuli. In the proposed study, verbal stimuli will be used as the visual component, which will have implications for the role of semantic awareness in the multisensory facilitation effect for verbal visual stimuli paired with congruent auditory stimuli. The proposed study is the first of its kind to examine the role of multisensory learning in the generation effect. For a summary of the study design and planned analyses see Table 1.

**References**

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14* (3), 257–262. <https://doi.org/10.1016/j.cub.2004.01.029>

Alais, D., Newell, F., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving, 23*(1), 3–38. <https://doi.org/10.1163/187847510X488603>

Barutchu, A., Freestone, D. R., Innes-Brown, H., Crewther, D. P., & Crewther, S. G. (2013). Evidence for enhanced multisensory facilitation with stimulus relevance: An electrophysiological investigation. *PLoS One, 8*(1), e52978. <https://doi.org/10.1371/journal.pone.0052978>

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. <https://doi.org/10.3758/bf03193441>

Burton, H., Snyder, A. Z., Diamond, J. B. and Raichle, M. E. (2002). Adaptive changes in early and late blind: A fMRI study of verb generation to heard nouns, *Journal of Neurophysiology 88*, 3359–3371. <https://doi.org/10.1152/jn.00129.2002>

Chang, A. C. S. (2009). Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories. *System, 37*(4), 652–663. <https://doi.org/10.1016/j.system.2009.09.009>

Diao, Y., & Sweller, J. (2007). Redundancy in foreign language reading comprehension instruction: Concurrent written and spoken presentations. *Learning and instruction, 17*(1), 78–88. <https://doi.org/10.1016/j.learninstruc.2006.11.007>

Ernst, M. and Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433. <https://doi.org/10.1038/415429a>

Fahey, S., Charette, L., Francis, C., & Zheng, Z. (2018). Multisensory integration of signals for bodily self-awareness requires minimal cognitive effort. *Canadian Journal of Experimental Psychology, 72*(4), 244–252. [https://doi.org/10.1037/cep0000152](https://psycnet.apa.org/doi/10.1037/cep0000152)

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002) Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience 22,* 5749–5759. <https://doi.org/10.1523/JNEUROSCI.22-13-05749.2002>

Fairhall, S. L., & Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience*, *29*(6), 1247–1257. <https://doi.org/10.1111/j.1460-9568.2009.06688.x>

Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods 41*, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research, 147*(3), 332–343. <https://doi.org/10.1007/s00221-002-1262-y>

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society, 360*, 815–836. <https://doi.org/10.1098/rstb.2005.1622>

Fuster, J.M. (1997). Network memory. *Trends in Neurosciences, 20*(10), 451–459. <https://doi.org/10.1016/S0166-2236(97)01128-4>

Gardiner, J. M., & Hampton, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 732–741. <https://doi.org/10.1037/0278-7393.11.1-4.732>

Ghazanfar, A.A., & Schroeder, C.E. (2006) Is neocortex essentially multisensory? *Trends in Cognitive Science 10*, 278–85. <https://doi.org/10.1016/j.tics.2006.04.008>

Gingras, G., Rowland, B. A., & Stein, B. E. (2009). The differing impact of multisensory and unisensory integration on behavior. *Journal of Neuroscience, 29*, 4897–4902.

Gottfried, J. A., Smith, A. P. R., Rugg, M. D. and Dolan, R. J. (2004). Remembrance of odors

past: human olfactory cortex in cross-modal recognition memory. *Neuron, 42*, 687–695. <https://doi.org/10.1016/S0896-6273(04)00270-3>

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425–2430. <https://10.0.4.102/science.1063736>

Heisz, J. J., Vakorin, V., Ross, B., Levine, B., & McIntosh, A. R. (2014). A trade-off between local and distributed information processing associated with remote episodic versus semantic memory. *Journal of Cognitive Neuroscience, 26*(1), 41–53. <https://doi.org/10.1162/jocn_a_00466>

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation.* Wiley, New York, 1966.

Kalyuga, S. (2000). When using sound with a text or picture is not beneficial for learning. *Australasian Journal of Educational Technology, 16*(2). <https://doi.org/10.14742/ajet.1829>

Kartal, G., & Simsek, H. (2017). The effects of audiobooks on EFL students’ listening comprehension. *The Reading Matrix: An International Online Journal, 17*(1), 112–123.

Kim, N. Y. (2021). E-Book, Audio-Book, or E-Audio-Book: The Effects of Multiple Modalities on EFL Comprehension. *English Teaching, 76*(4), 33–52.

Kim RS, Seitz AR, Shams L (2008) Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS ONE 3*(1), e1532. <https://doi.org/010.1371/journal.pone.0001532>

Kinoshita, S. (1989). Generation enhances semantic processing? The role of distinctiveness in the generation effect. *Memory & Cognition*, *17*(5), 563–571. <https://doi.org/10.3758/bf03197079>

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory–visual integration. *Behavioural Brain Research, 121*(1–2), 69–79. <https://doi.org/10.1016/s0166-4328(00)00386-7>

Liu, H., Cao, S., & Wu, S. (2019). An experimental comparison on reading comprehension effect of visual, audio and dual channels. *Proceedings of the Association for Information Science and Technology, 56*(1), 716–718. <https://doi.org/10.1002/pra2.148>

Mayer, R. E. (2001). *Multimedia Learning.* New York: Cambridge University Press.

McCurdy, M. P., Viechtbauer, W., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (2020). Theories of the generation effect and the impact of generation constraint: A meta-analytic review. *Psychonomic Bulletin & Review, 27,*1139–1165. <https://doi.org/10.3758/s13423-020-01762-3>

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407*(6806), 906–908. <https://doi.org/10.1038/35038085>

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748. <https://doi.org/10.1038/264746a0>

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition.* Psychology Press.

Meredith, M. A., Wallace, M. T., & Stein, B. E. (1992). Visual, auditory and somatosensory convergence in output neurons of the cat superior colliculus: Multisensory properties of the tecto-reticulo-spinal projection. *Experimental Brain Research, 88*(1), 181–186. <https://doi.org/10.1007/bf02259139>

Mohamed, M. M. K. (2018). Using audiobooks for developing listening comprehension among Saudi EFL preparatory year students. *Journal of Language Teaching and Research, 9*(1), 64–73. <http://dx.doi.org/10.17507/jltr.0901.08>

Morais, A. S., Olsson, H., & Schooler, L. J. (2013). Mapping the structure of semantic memory. *Cognitive Science, 37*(1), 125–145. <https://doi.org/10.1111/cogs.12013>

Moran, Z. D., Bachman, P., Pham, P., Cho, S. H., Cannon, T. D., & Shams, L. (2013). Multisensory encoding improves auditory recognition. *Multisensory Research, 26*(6), 581–592. <https://doi.org/10.1163/22134808-00002436>

Nyberg, L., Habib, R., McIntosh, A. R. and Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval, *Proceedings of the National Academy of Sciences, 97*, 11120–11124. <https://doi.org/10.1073/pnas.97.20.11120>

Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man; a clinical study of localization of function.* Macmillan.

Plass, J. L., Moreno, R., & Brünken, R. (Eds.). (2010). Cognitive load theory.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects, *Nature Neuroscience, 2*, 79–87. <https://doi.org/10.1038/4580>

Rockland K.S., Ojima, H. (2003) Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology 50,* 19–26. <https://doi.org/10.1016/S0167-8760(03)00121-1>

Röder, B., Stock, O., Bien, S., Neville, H. and Rosler, F. (2002). Speech processing activates visual cortex in congenitally blind humans, *European Journal of Neuroscience 16*, 930–936. <https://doi.org/10.1046/j.1460-9568.2002.02147.x>

Rogowsky, B. A., Calhoun, B. M., & Tallal, P. (2016). Does modality matter? The effects of reading, listening, and dual modality on comprehension. *Sage Open, 6*(3). <https://doi.org/10.1177/2158244016669550>

Rolls, E. T., Treves, A. and Tovee, M. J. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Experimental Brain Research, 114*, 149–162. <https://doi.org/10.1007/PL00005615>

Seitz, A., Dinse, H. (2007) A common framework for perceptual learning. *Current Opinion of Neurobiology 17*(2), 148–153. <https://doi.org/10.1016/j.conb.2007.02.004>

Seitz, A.R., Kim, R., & Shams, L. (2006) Sound facilitates visual learning. *Current Biology, 16*, 1422–1427. <https://doi.org/10.1016/j.cub.2006.05.048>

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*(11), 411–417. <https://doi.org/10.1016/j.tics.2008.07.006>

Shams, L., Wozny, D. R., Kim, R. & Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Frontiers in Psychology, 2*, 264. <https://doi.org/10.3389/fpsyg.2011.00264>

Sheffert, S. M., & Olson, E. (2004). Audiovisual speech facilitates voice learning. *Perception in* *Psychophysics, 66*, 352–362. <https://doi.org/10.3758/BF03194884>

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, *4*(6), 592–604. <https://doi.org/10.1037/0278-7393.4.6.592>

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport, 12*(1), 7–10. <https://doi.org/10.1097/00001756-200101220-00009>

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1517–1529. <https://doi.org/10.1037/a0027339>

Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *The Quarterly Journal of Experimental Psychology Section A, 36*(1), 51–74. <https://doi.org/10.1080/14640748408401503>

Treisman, A. M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136. <https://doi.org/10.1016/0010-0285(80)90005-5>

Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature, 383*(6597), 254–256. <https://doi.org/10.1038/383254a0>

Von Kriegstein, K. & Giraud, A. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology, 4*, 1809–1820. <https://doi.org/10.1371/journal.pbio.0040326>

Wallace, M., Roberson, G., Hairston, W., Stein, B., Vaughan, J., & Schirillo, J. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research, 158*(2). <https://doi.org/10.1007/s00221-004-1899-9>

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia, 64*, 105–123. <https://doi.org/10.1016/j.neuropsychologia.2014.08.005>

Wallace, M. T., Wilkinson, L. K., & Stein, B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology, 76*(2), 1246–1266. <https://doi.org/10.1152/jn.1996.76.2.1246>

Wheeler, M. E., Petersen, S. E. and Buckner, R. L. (2000). Memory’s echo: vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences, 97*, 11125–11129. <https://doi.org/10.1073/pnas.97.20.11125>

Wyk, B. C., Ramsay, G. J., Hudac, C. M., Jones, W., Lin, D., Klin, A., Lee, S. M., & Pelphrey, K. A. (2010). Cortical integration of audio–visual speech and non-speech stimuli. *Brain and Cognition, 74*(2), 97–106. <https://doi.org/10.1016/j.bandc.2010.07.002>

**Table 1**

*Study Design*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Question** | **Hypothesis**  (broken down in manuscript under “Hypotheses”) | **Sampling plan** | **Analysis Plan** | **Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis** | **Interpretation given different outcomes** | **Theory that could be shown wrong by the outcomes** |
| How does the presence of multisensory facilitation influence the generation effect in memory tasks, particularly in terms of subsequent recognition of target items? | H1: Multisensory > unisensory in enhancing recognition of target items during generation tasks. Specifically, audiovisual > auditory or visual conditions | A sample size of 72 participants was determined using G\*Power to achieve a power of 0.90 to detect a medium effect size, consistent with effect sizes observed in past work. | A 2 (Task Type: generate, read) x 3 (Sensory Modality: auditory, visual, audiovisual) factorial ANOVA will be conducted to assess the effects of task type and sensory modality on recognition outcomes. Confidence ratings will be analyzed using independent samples t-tests to compare task types. Bayes’ analysis will follow up n.s. findings. | Effect size determined based on prior research suggesting a medium effect size for the hypothesized differences in recognition outcomes for such tasks. | **H1a**: A significant Sensory Modality effect in the ANOVA, with highest recognition in audiovisual condition, supports multisensory > unisensory in enhancing target item recognition.  **H1b**: A significant Task Type × Sensory Modality interaction in ANOVA, showing audiovisual > auditory or visual conditions specifically in generation task, further supports audiovisual superiority. | Theoretical framework: Explores how the generation effect and multisensory facilitation impact memory.  Generation effect: self-generated info > passive reading (Slamecka & Graf, 1978).  Multisensory facilitation: multiple senses > better learning (Shams & Seitz, 2008).  Potential Challenge to Existing Theories:  Early notion: auditory-visual pairing boosts visual processing (Seitz et al., 2006).  If results show facilitation only in congruent conditions, alerting effect may be wrong.  Integration of congruent info enriches memory, challenging redundancy theory (Diao & Sweller, 2007; Moran et al., 2013). Integration of congruent info may not enhance memory, supporting redundancy theory (Diao & Sweller, 2007; Moran et al., 2013). |
| What is the impact of task type (reading vs. generation) on recognition outcomes and confidence levels, and how does this vary across different sensory conditions? | H2: Generation > reading tasks in recognition outcomes and confidence. Moreover, audiovisual generation > unisensory reading, with a greater discrepancy in outcomes. | See above. | See above. |  | **H2a**: A significant Task Type effect in ANOVA, indicating higher recognition in generation than reading tasks, supports generation > reading in recognition outcomes.  **H2b:** A significant Task Type × Sensory Modality interaction, with greater discrepancy in audiovisual generation vs. unisensory reading, supports audiovisual generation > unisensory reading tasks.  **For both hypotheses:**  Bayes' analysis providing evidence for null hypothesis undermines proposed hypotheses, while evidence against null strengthens confidence in hypotheses. |

**Appendix**

**List of Word Pairs**

1. Woods-forest
2. Pleasant-nice
3. Require-need
4. Wealthy-rich
5. Sad-upset
6. Cash-money
7. Chilly-cold
8. Shack-cabin
9. Wail-weep
10. Terminate-finish
11. Filthy-dirty
12. Identical-same
13. War-battle
14. Safe-secure
15. Start-begin
16. Shove-push
17. Ship-boat
18. Cab-taxi
19. Woman-lady
20. Throw-toss
21. Employ-hire
22. Mad-angry
23. Advance-progress
24. Informal-casual
25. Attain-achieve
26. Intent-purpose
27. Elaborate-complex
28. Rescue-save
29. Advertise-promote
30. Difficult-hard
31. Water - Fluid
32. River - Stream
33. Clock - Timer
34. Car-vehicle
35. Beverage - Drink
36. Chuckle - Laugh
37. Waltz - Dance
38. House - Home
39. Smile - Grin
40. Fashion - Style
41. Earth - World
42. Pony - Horse
43. Sound - Noise
44. Furnace-heater
45. Night - Evening
46. Drizzle - Rain
47. Amazing – Incredible
48. Liquor-alcohol
49. Huge-Enormous
50. Radiant-Bright
51. Tranquil-Peaceful
52. Carve-Slice
53. Danger-hazard
54. Idea-thought
55. Help-assist
56. Murder-kill
57. Book-novel
58. Love-admire
59. Peculiar-unusual
60. Rabbit-bunny
61. Instructor-teacher
62. Little-small
63. Despise-hate
64. Under-below
65. Connect-link
66. Physician-doctor
67. Grass-lawn
68. Mistake-error
69. Enemy-opponent
70. Harm-damage
71. Infant-baby
72. Depart-leave
73. Robber-thief
74. Paste-glue
75. Drowsy-sleepy
76. Quiet-silent
77. Steer-drive
78. Come-arrive
79. Lose-fail
80. Expand-stretch
81. Delicate-fragile
82. Untrue-False
83. Question-inquiry
84. Admit-confess
85. Vacant-empty
86. Close-shut
87. Broad-wide
88. Respond-Answer
89. Stop-quit
90. Gift-present
91. Above-over
92. Occur-happen
93. Illness-disease
94. Scream-yell
95. Stone-rock
96. Approve-accept
97. Illicit-illegal
98. Humble-modest
99. Plausible-possible
100. Gain-acquire
101. Hefty-heavy
102. Thankful-grateful
103. cheerful-joyful
104. Agitated-annoyed
105. Unsure-doubtful
106. Cautious-careful
107. Vulnerable-sensitive
108. Blame-accuse
109. Combat-fight
110. Divorce – separate
111. Comprehend-understand
112. Obligate-force
113. Artificial-fake
114. Beautiful-gorgeous
115. Barbecue-grill
116. Waiter-server
117. Spouse-partner
118. Duvet-comforter
119. Priest-pastor
120. Speak-talk