Visual and Auditory Stimuli Pairings During Recognition Memory

Grace Liu

Michael Miller, Ph.D., Evan Layher, Ph.D.

Department of Psychological & Brain Sciences, University of California, Santa Barbara

Introduction

- Dual-coding theory: information is encoded via two channels – verbal and nonverbal (Clark & Paivio, 1991). Pairing a visual image with an auditory cue can create a strong cognitive bridge in associative memory (Paivio, 1969).
- Previous research found that auditory-visual interactions occur early in perceptual processing in the primary visual cortex, which was previously thought to be only responsible for visual information (Thesen et al., 2004). Visual stimuli occurring with sounds enhanced the brain's response to sounds.
- Auditory recognition memory is weaker than visual recognition memory (Cohen et al., 2009).
- Signal detection theory (SDT): analyzes decision-making performance when interference is present (Swets et al., 1961), showing how one discriminates between signal vs. noise assuming neither is perfectly independent.
- Measuring recognition accuracy across visual/auditory stimuli pairings can reveal how multisensory experiences may affect recognition memory, aiding research in recognition memory strategies, education, and memory-related disorders.
- The study aims to discern if familiar/unfamiliar visual stimuli paired with familiar/unfamiliar auditory stimuli can improve recognition, creating a memory recognition hierarchy based on stimulus familiarity.

Hypotheses

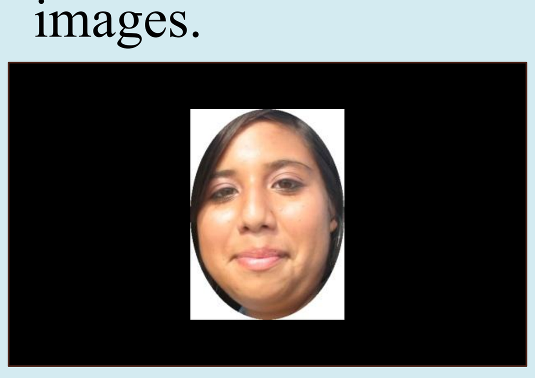
H1: Pairing visual face images with sounds will lead to more robust memory encoding than exposure to only one sense stimuli.
H2: Pairing familiar faces with sounds will lead to higher recognition accuracy for the sounds than pairing unfamiliar faces with sounds.
H3: Recognition accuracy for faces will be higher than audios across all three conditions.

Methods

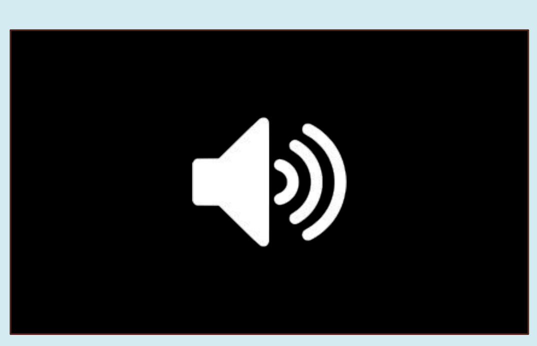
Participants: 72 participants (53 = female, 17 = male, 2 = non-binary) between the ages of 18 and 30 (*M* = 20.44 , *SD* = 1.943), recruited from UCSB SONA.

  - 5 participants were excluded due to below chance performance.

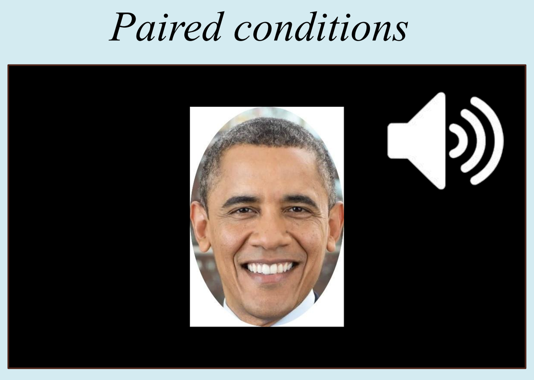
- Participants were exposed to six study/test phase cycles with visual and auditory stimuli pairing conditions and asked to recognize a set of stimuli.
- **Conditions:** [1] visual stimuli only [2] visual stimuli with unfamiliar audio [3] visual stimuli with familiar audio [4] auditory stimuli only [5] auditory stimuli with unfamiliar images [6] auditory stimuli with familiar images.



Condition [1]

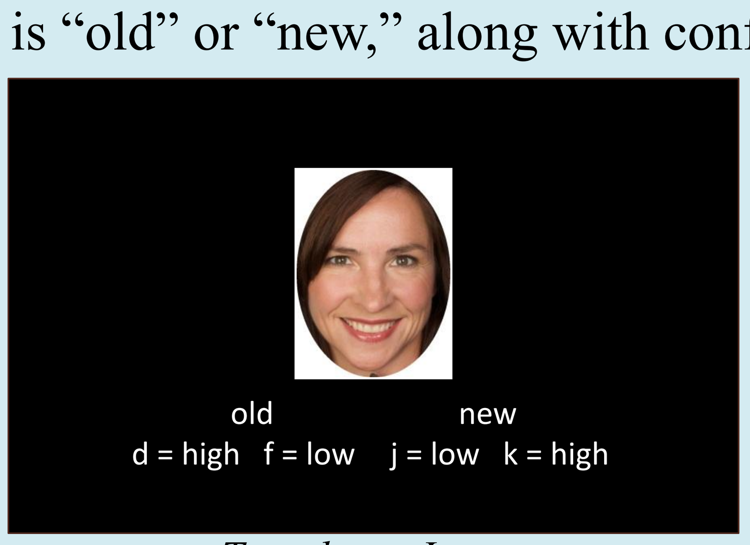


Condition [4]

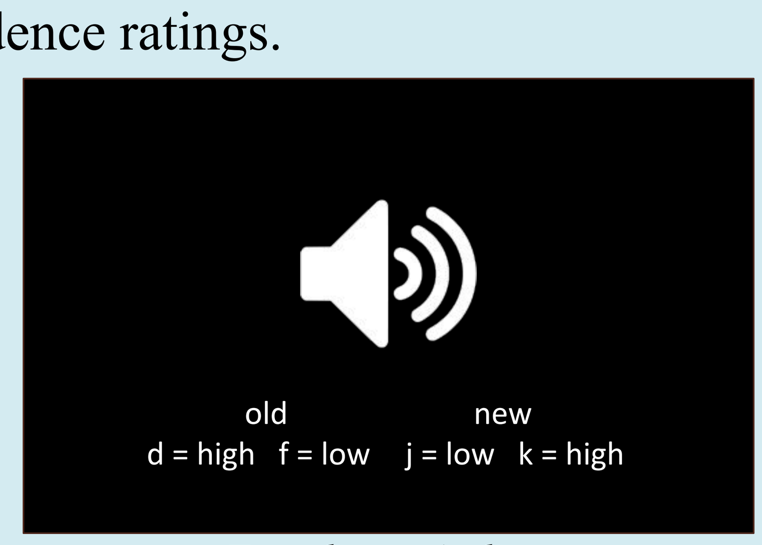


Paired conditions

- Test phases presented participants with a studied image/audio or a foil image/audio, and are then asked to determine whether the image/audio is "old" or "new," along with confidence ratings.



Test phase: Image



Test phase: Audio

- A SDT curve measures how well signal vs. noise (familiar vs. unfamiliar stimuli) is discriminated (d'). A greater d' means a person can more accurately differentiate between the two (Banks, 1970).

Results

- Of the six conditions, there was only a significant effect on recognition memory in condition [1] ($p = 0.0238$).
- Condition [1] had the greatest d' (1.829).
  - The visual stimuli condition had a greater average d' (1.678) than the auditory stimuli condition d' (0.743) (Figure 1).

Condition	Test	Pair	d'
[1]	Image	No audio	1.829
[2]	Image	Unfamiliar audio	1.615
[3]	Image	Familiar audio	1.589
[4]	Audio	No image	0.767
[5]	Audio	Unfamiliar image	0.680
[6]	Audio	Familiar image	0.781

Figure 1. Mean d' values of each condition

- The ROC curve illustrates participants' recognition accuracy performance across all six conditions at different confidence ratings.

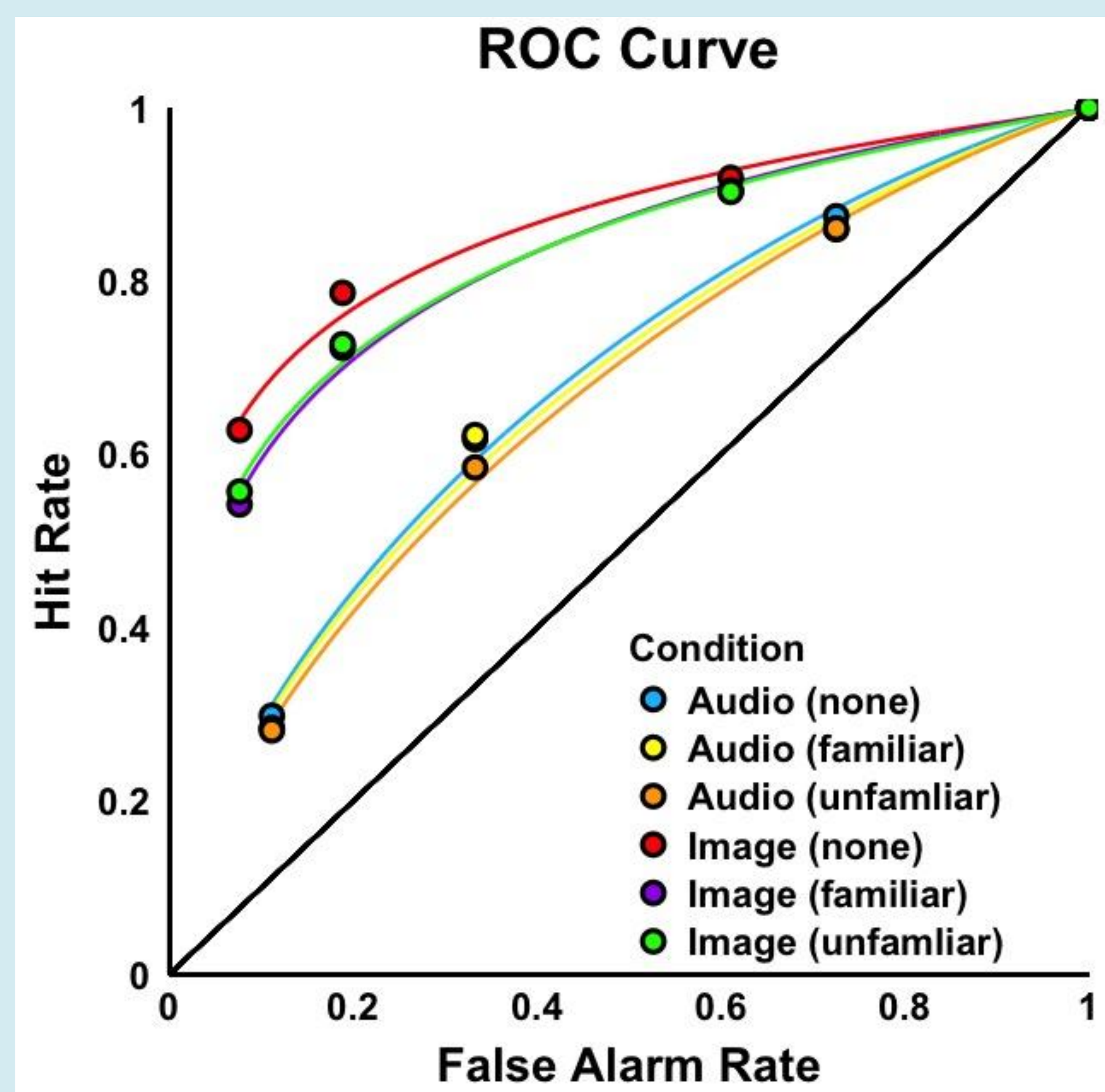


Figure 2.

Discussion

- **H1**: Pairing visual face images with sounds did not improve memory performance, but instead decreased visual stimuli performance.
- **H2**: Pairing familiar faces with sounds had no significant effect on recognition accuracy for the sounds compared to pairing unfamiliar faces with sounds.
  - Conditions that had a pairing, instead of just one stimuli, had lower recognition accuracy than conditions with no pairing.
- **H3**: Recognition memory for images is stronger than recognition memory for audios, supporting previous findings.

References

1. Banks, W. P. (1970). Signal detection theory and human memory. Psychological bulletin, 74(2), 81.
2. Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. Educational psychology review, 3, 149-210.
3. Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. Proceedings of the National Academy of Sciences, 106(14), 6008-6010.
4. Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological review*, *76*(3), 241.
5. Swets, J. A., Tanner Jr, W. P., & Birdsall, T. G. (1961). Decision processes in perception. Psychological review, 68(5), 301.
6. Thesen, T., Vibell, J. F., Calvert, G. A., & Österbauer, R. A. (2004). Neuroimaging of multisensory processing in vision, audition, touch, and olfaction. Cognitive Processing, 5, 84-93.

Acknowledgements

I would like to thank Evan Layher for his guidance, wisdom, and support on this project. I would also like to thank Dr. Michael B. Miller and the Miller Lab for their support throughout my project. *Sponsored by the Undergraduate Research Creative Activities Grant (URCA).*

Pairing Visual and Auditory Stimuli During Recognition Memory

Grace Liu, Evan Layher, Ph.D., and Michael Miller, Ph.D.

Department of Psychological & Brain Sciences

University of California, Santa Barbara

Abstract

For many students, we endeavor to find the best study methods for remembering information. We try reading notes, listening to lectures, or doing both simultaneously. However, which is the best method for retaining it all? The present study aims to determine how multisensory experiences can affect recognition memory. During the study phase, participants were exposed to simultaneous visual and auditory pairings in which unfamiliar stimuli in one modality were presented with stimuli from the other modality that were either [1] unfamiliar, [2] familiar, or [3] no stimuli at all (control condition). At test phases, participants viewed either studied or novel stimuli from one modality and needed to determine if they remembered each stimulus. Overall, it was found that visual recognition memory exceeds auditory recognition memory, but there was no significant effect of the stimulus pairing status (familiar vs. unfamiliar) from the other sensory modality. In fact, we found that recognition memory performance was only significantly improved when participants studied visual stimuli only. This suggests that presenting auditory stimuli during the study phase actually hinders recognition performance for visual stimuli. Together, these findings indicate that presenting visual and auditory stimuli together does *not* improve recognition memory performance of individual stimuli.

Introduction

Our ability to recognize and retrieve information from our memory, the foundation of human cognition, is something that most people may not consciously consider or think about day-to-day. Such information stems from our experiences, gathered through our five senses: sight, sound, taste, smell, and touch. None of these senses are processed individually; crossmodal perceptions have been established in perceptual processing, and these interactions between two or more senses improve the precision, accuracy, and processing speed of perception (Shams et al., 2011).

Since multisensory processing requires increased activation in the brain, does that mean that things that were exposed through more than one sense pathway are retained in our memory better?

Multisensory stimuli processing

According to the dual-coding theory (Clark & Paivio, 1991), information is processed and encoded in the mind along two different channels: verbal and nonverbal. There is plenty of research surrounding auditory-visual (AV) interactions in perceptual processing. In auditory processing, studies have shown that the primary auditory cortex is influenced by visual stimuli. Through neuroimaging techniques, AV multisensory interactions were found to occur early in processing in the primary visual cortex, which was previously thought to be only responsible for visual information (Thesen et al., 2004). Thus, visual stimuli occurring with sounds enhance the brain's response to sounds. This differs from previous ideas that the brain processes visual and auditory stimuli separately and then combines them later (Engen et al., 2002). In a study on non-human primates by Falchier et al. (2002), early auditory influence was found on visual processing where direct neuronal connections were found between the auditory and primary visual cortexes. Another study with rats found that visual cues enhanced primary auditory cortex neural responses, aiding in the discrimination of different sounds (Chang et al., 2022). Such investigations illustrate how sight and sound are not processed independently, but processed together to enhance sensory perception.

Facial visual processing areas

The face is the first encounter with someone's identity and, thus, numerous brain areas are allocated for face processing. These areas lie in the occipitotemporal cortex and, through the use of functional magnetic resonance imaging (fMRI), it has been found that the fusiform face area (FFA), located in the right fusiform gyrus of the primary visual cortex (V1), is active when looking at faces (Kanwisher et al., 1997). Other core face-selective areas include the posterior superior temporal sulcus (pSTS) and the occipital face area (OFA) (Nikel et al., 2002). Within V1, the fusiform gyrus is located at the lower surface of the occipital and temporal lobes and plays a significant role in visual processing (Weiner & Zilles, 2016). It is the largest structure and has many different functions including object recognition, face and body perception, and processing visual stimuli, among others. Within the fusiform gyrus, the FFA has been identified as the key region in processing facial identity (Grill-Spector et al., 2004). The fMRI is the preferred method for imaging brain areas and their activation due to its high spatial resolution that allows us to follow the oxygenation level-dependent (BOLD) signals that occur during neuronal activation. Previous studies have established FFA activation in both facial detection and identification through image identification tasks and analyzing BOLD signals provided by fMRI.

Auditory processing areas

Auditory processing also involves multiple specialized areas, located in the temporal lobes. The auditory cortex is essential for processing all types of sounds, and recognizing and differentiating them. It mainly consists of the posterior side of the superior temporal gyrus (STG), including Heschl's gyrus (HG) and Planum Temporalte (PT), and is the central region for auditory processing (Trébuchon et al., 2021). The primary auditory cortex processes a variety of sound features, such as loudness, tone, and pitch (Morillon et al., 2002).

Recognition memory

Auditory recognition memory is inferior to visual recognition memory for individual items (Cohen et al., 2009). However, when items are presented together, associations between stimuli form, which is known as associative recognition memory (Ngo, 2013). Associative memory could establish a strong cognitive bridge by associating a visual image with an auditory cue (Paivio, 1969). Further, familiarity and novelty in stimuli should also be considered. A previous study in recognition memory for faces has found that familiarity can support associative recognition judgments (Yonelinas, 1999). This means that being familiar with something can strengthen your associative memory, or strengthen your memory for multiple stimuli. This is in line with the mere exposure effect, which explains how there is a preference for something that has been encountered before, or something familiar (Zajonc, 1968). Thus, this implies that if there are familiar stimuli present in a stimuli pairing, then the recognition will be stronger for the pairing.

Signal detection theory

The signal detection theory (SDT) analyzes decision-making performance when interference is present (Swets et al., 1961). More specifically, the SDT helps understand how a person discriminates between a signal vs. noise with the assumption that neither is perfectly independent. This is done by placing individuals in scenarios where they must identify between two possible options (Lloyd & Appel, 1976). In the context of the present study, the signal is the familiar stimuli and the noise is the unfamiliar stimuli. In SDT, there are four outcomes when decisions are made:

1. Hit: correctly identifies a familiar stimulus
2. Miss: incorrectly rejects a familiar stimulus
3. False alarm (FA): incorrectly accepts an unfamiliar stimulus as a familiar stimulus
4. Correct rejection (CR): correctly rejects an unfamiliar stimulus

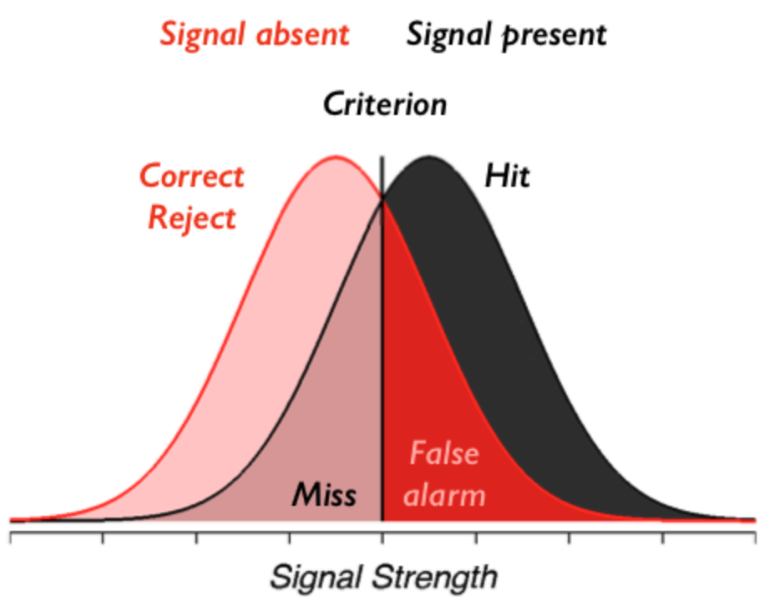


Figure 1. Signal Detection Theory graph

The main measurement of the SDT curve is *d'*, which is how well a person can discriminate between signal and noise (or familiar and unfamiliar stimuli). A greater *d'* means a person can more accurately differentiate between the two and is represented by the distance between the mean of the signal present versus absent distributions (Banks, 1970). Response biases, a person's tendency to report the presence of a signal in an ambiguous situation, are measured through criterion placement (*c*) (Macmillan & Creelman, 2005). A liberal bias (negative *c*) occurs when a person has a lower threshold for deciding that a signal is present, often leading to more hits, but also more FA (prefer to risk FA than miss a signal). A conservative bias (positive *c*) occurs when a person has a higher threshold for deciding that a signal is present, often leading to fewer hits and, therefore more misses, but fewer false alarms — essentially playing it safe. Thus, *c* quantifies if a person is more cautious or not with their decisions.

Present study

Based on previous research, we propose that presenting visual and auditory stimuli simultaneously will activate multiple sensory information encoding channels, leading to improved recognition memory of individual items. While individuals are expected to have greater recognition performance for visual versus auditory stimuli, we predict both will be enhanced when another stimulus from the other modality is presented simultaneously. Further, within the stimuli pairings types (familiar vs. unfamiliar), individuals are expected to recognize stimuli that were paired with familiar stimuli better, in accordance with the mere exposure theory.

The study aims to discern whether familiar or unfamiliar visual stimuli paired with familiar or unfamiliar auditory stimuli can improve recognition, and thus create a hierarchy of memory recognition based on stimulus familiarity. Understanding these dynamics could provide insights into cognitive processes involved in encoding and retrieving memories, and could contribute to educational, clinical, and technological applications.

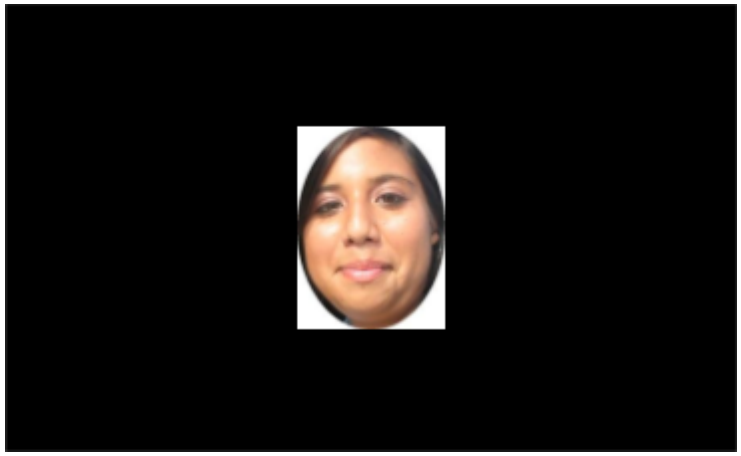
The present study includes 1 experiment, with six stimulus pairing conditions. From this task, the results of each condition are compared to determine the greatest recognition memory performance.

Methods

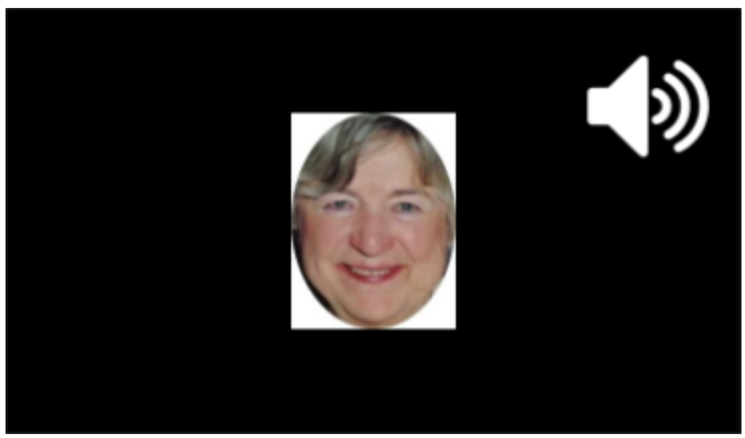
Design

The experiment design consisted of a 2-sensory input X 3-stimulus pairing experimental design, creating six test conditions. Participants underwent six study/test phase cycles where they were exposed to visual and auditory stimuli conditions and asked to memorize a set of stimuli — some study phase blocks will include pairings with familiar/unfamiliar/no stimuli (refer to Appendix A for a list of familiar stimuli used). The six test conditions are as follows:

[1] visual stimuli that was not presented with any audio during the study phase



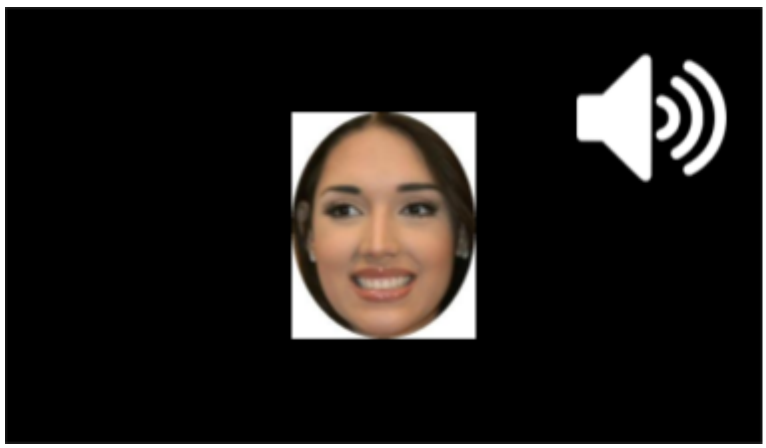
[2] visual stimuli presented with unfamiliar audio during the study phase



[3] visual stimuli presented with familiar audio during the study phase

[4] auditory stimuli that was not presented with any image during the study phase

[5] auditory stimuli presented with unfamiliar images during the study phase



[6] auditory stimuli presented with familiar images during the study phase

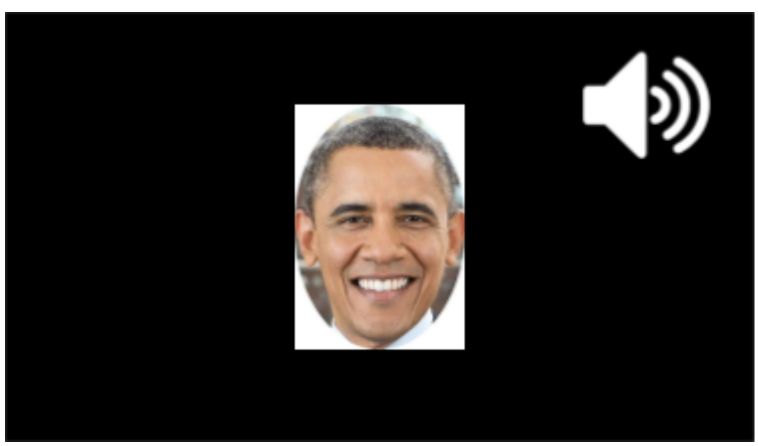


Figure 2. Six test conditions formatting

Still, there are only five unique study phase conditions as the unfamiliar visual stimuli with unfamiliar auditory stimuli condition will be conducted twice (once for condition [2] and another for condition [5] listed above). Half of the test stimuli are previously studied visual and auditory stimuli and the other half are novel visual and auditory stimuli. This experiment allowed us to observe the effect of familiar/unfamiliar sense stimuli pairings on decision-making and recognition memory.

Participants

There were 72 participants (53 = females, 17 = males, 2 = non-binary) between the ages of 18 and 30 years old (*M* = 20.44, SD = 1.943). Participants were recruited via SONA from the University of California, Santa Barbara (UCSB) undergraduate population. 5 participants were excluded from the data analysis as their results indicated that they were answering randomly or at below chance performance (this was determined if their *d'* was less than 0.5). Participants received $10 in cash in compensation, funded by the Undergraduate Research & Creative Activities (URCA) Grant.

Materials

The computer task was employed synchronously, in-person at the lab, utilizing Psychtoolbox in MATLAB. Participants were first presented with a consent form using Qualtrics that they had to respond to before beginning the task. Within the task, participants answer questions regarding participant demographics, such as age, gender, race, and ethnicity. Task instructions were then presented to participants. Responses during the study and test phases were digitally recorded.

A total of 36 familiar face stimuli were found and edited using Google Images and macOS Preview. A total of 36 familiar audio stimuli were found using Apple Music and YouTube, and then recorded and edited with Apple Voice Memos. A total of 216 unfamiliar face stimuli were found using a public database (Bainbridge et al., 2013). A total of 216 unfamiliar audio stimuli were found using the AI music generator Soundraw, and then recorded and edited with Apple Voice Memos. All face stimuli consisted of diverse ethnicities and genders; familiar face stimuli consisted of a range of famous people, such as Barack Obama, Britney Spears, and Lionel Messi. Familiar audio stimuli consisted of audios that are generally well-known internationally, including "Happy Birthday," "Fur Elise," and common alarm sounds. Unfamiliar audio stimuli were tunes that ranged in genre, from orchestral to electronic dance music.

During the test phases, each face and audio stimuli were rated by the participant using four keys: d = "high" confident "old," f = "low" confident "old," j = "low" confident "new," and k = "high" confident "new." This allowed participants to gauge their confidence in recognizing familiar and unfamiliar visual and audio stimuli.

Procedure

When participants arrived at the laboratory, they were instructed to complete a consent form through Qualtrics and were to decide whether or not to proceed. Participants were then instructed to begin the task in MATLAB and respond to four demographic questions regarding age, gender, race, and ethnicity. Following that, participants were presented with the instructions; they were told to press one of the following four keys to make a response:

d = “high” confident “old”

f = “low” confident “old”

j = “low” confident “new”

k = “high’ confident “new”

Test phase: Image *Test phase: Audio*

Figure 3. Test phase formatting

Participants were directed through a practice cycle of a study phase, followed by a test phase. Throughout the study phase, the participants were shown visual/auditory stimuli that could be familiar or unfamiliar. The study phase was then followed by a memory test in which participants were presented with a studied image/sound or a foil image/sound, and were asked to determine whether the image/sound was “old” or “new,” along with their confidence rating. Each test phase has a study phase that meets the test phase conditions. The participants then proceeded to complete six official study/test phase cycles. The computer task took approximately 45 minutes to complete.

Calculations

Signal Detection Theory

As mentioned previously in the introduction, this study relies on SDT to calculate the discriminability in recognizing stimuli (*d'*) and response biases (*c*). For each of the six conditions, the hit and false alarm rates are used to calculate *d'*. The hit and false alarm rates are also used to create an ROC curve, which depicts the performance of all conditions, at varying confidence levels. The ROC curve is a method of visualizing the SDT.

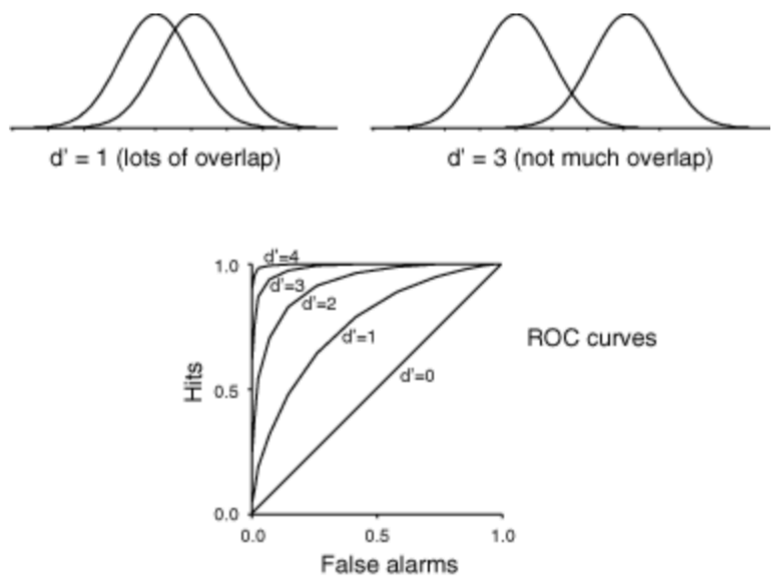


Figure 4. d' plotted as a ROC curve.

<u>Results</u>

We hypothesized that pairing visual face images with sounds during initial exposure will lead to better recognition memory performance than exposure to only one sense stimuli during initial exposure, while also considering the use of familiar and unfamiliar stimuli. To analyze how well participants recognized stimuli during the test phases for all six conditions, we calculated the mean *d'* of each condition and conducted ANOVAs across image and audio conditions.

Recognition memory performance (*d'*)

The SDT revealed that across all visual stimuli conditions, there was a greater average *d'* (*d'* = 1.678) than across all auditory stimuli conditions (*d'* = 0.473) (Figure 5). Image-only and audio-only ANOVAs found that there was only a significant effect in the image-only condition (p = 0.0238), while there was no significant effect in the audio-only condition (p = 0.221).

Condition	**Test**	**Pair**	***d'***	**SD**
[1]	Image	No audio	1.829	0.402
[2]	Image	Unfamiliar audio	1.615	0.373
[3]	Image	Familiar audio	1.589	0.336
[4]	Audio	No image	0.767	0.351
[5]	Audio	Unfamiliar image	0.680	0.353
[6]	Audio	Familiar image	0.781	0.338

Figure 5. Mean d' values of each condition.

Response biases (c)

The participants' decision threshold, or criterion placement, varied for each condition (Figure 6). The *c* was substantially smaller for condition [1] ($c = 0.028$) than all other conditions. The average *c* for the image and audio conditions are $c = 0.104$ and $c = 0.082$, respectively.

Condition	**Test**	**Pair**	***c***	**SD**
[1]	Image	No audio	0.028	0.196
[2]	Image	Unfamiliar audio	0.135	0.187
[3]	Image	Familiar audio	0.148	0.197
[4]	Audio	No image	0.069	0.174
[5]	Audio	Unfamiliar image	0.114	0.196
[6]	Audio	Familiar image	0.063	0.174

Figure 6. Mean c values of each condition.

To observe the recognition accuracy performance across all six conditions at different confidence ratings, a ROC curve was constructed (Figure 7). Each condition is represented by four points, corresponding to different levels of confidence: high confidence for old stimuli, low confidence for old stimuli, low confidence for new stimuli, high confidence for new stimuli (from left to right). At both the high and low confidence levels in old stimuli for both visual and auditory stimuli, there are lower FA rates compared to both the high and low confidence levels for new stimuli. Condition [1] (red) has the highest hit rates across all confidence levels.

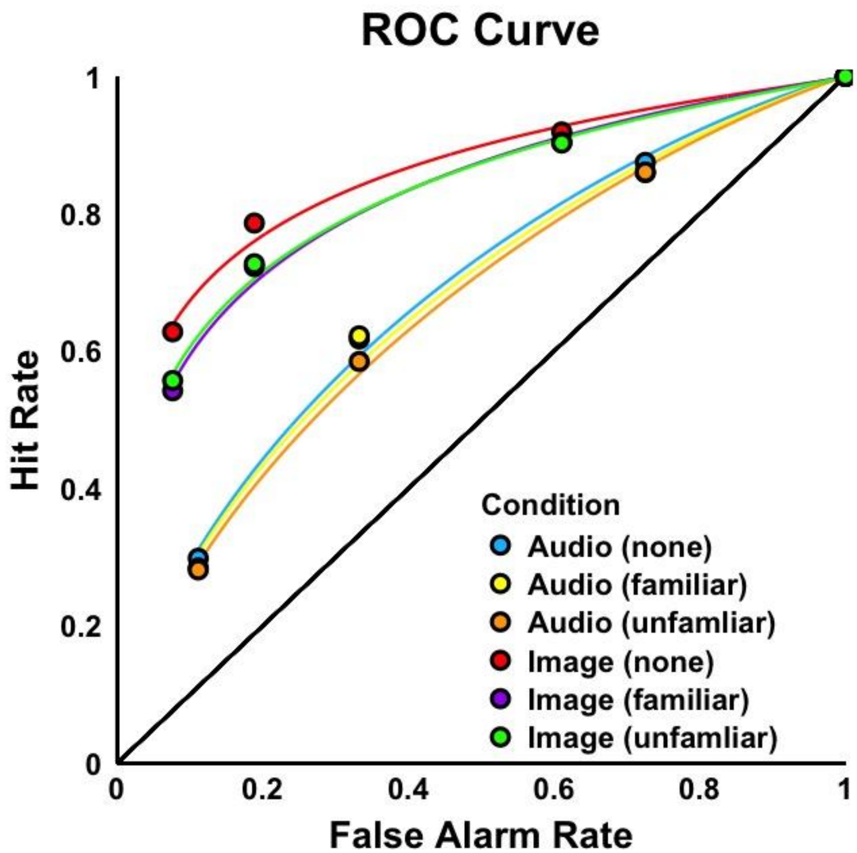


Figure 7. ROC curve for d' of each of the 6 test conditions. Each color represents the stimulus type (audio vs. image) and the stimulus type that it is paired with (no stimulus, familiar stimulus, unfamiliar stimulus). The four points along each of the 6 curves represent the four different confidence levels.

Discussion

This study hypothesized that exposure to stimuli through multiple senses at once would strengthen the encoding of these stimuli, and would therefore be better recognized later. However, it was found that pairing visual face images with sounds did not improve memory performance. By looking at the *d'*, which measured how well participants were able to recognize stimuli that they were exposed to, it was found that recognition for the visual stimuli actually decreased when an auditory stimulus was also presented. It is possible that the auditory stimulus actually interfered with the encoding of the visual stimuli, leading to worse performance. Memory for auditory stimuli was not significantly affected by the presence of an image at encoding.

Stimuli pairings differed in the stimuli familiarity, as it was hypothesized that stimuli that were paired with familiar stimuli would be remembered more. It was found that familiarity with stimuli had no significant effect on stimuli recognition. Participants performed about the same regardless of whether the stimuli were accompanied by a familiar face, an unfamiliar face, a familiar sound, or an unfamiliar sound at first exposure. In fact, conditions that did have a pairing (where a visual stimulus was paired with an auditory stimulus), had lower recognition accuracy than conditions with no pairing in the visual stimuli conditions. Although the audio pairing with familiar visual stimuli had the greatest recognition accuracy among the three auditory stimuli conditions, exceeding the audio-only condition, there is no significant difference. In the end, it seems that participants' recognition memory was greatest for stimuli that they had initially been exposed to alone.

In line with previous research where it was found that recognition memory is better for visuals than audio (Cohen et al., 2009), the present study determined that recognition memory for visual stimuli was better than recognition memory for auditory stimuli. This is also likely why participants were on average more liberal with their decision-making when identifying audio stimuli than visual stimuli, as depicted in the SDT *c* calculations. Since the auditory stimuli were harder to identify, participants opted to take the chance/take the risk that the stimulus was one they had encountered before. This can be supported by the findings of Botta et al. (2011), which suggested that auditory stimuli are associated with higher difficulty and lower confidence. However, it is surprising that participants were extremely liberal when it came to identifying visual stimuli that they were exposed to alone. It would have been expected that participants would be more conservative, or not take the risk, when identifying these faces as this condition had the greatest recognition memory accuracy.

The study sought to determine if pairing visual and auditory stimuli together during encoding could improve memory performance of individual items, particularly when paired with familiar items. While previous studies have suggested that integrating information across multiple senses could improve learning and retention (Shams & Seitz, 2008), the present study's findings indicate that this is not always the case, especially when it comes to recognition memory for individual items. Results showed that memory for visual stimuli was optimal when no audio was presented with it, suggesting that simultaneous presentation of audio actually disrupts encoding and subsequent memory of visual stimuli rather than facilitating it. For auditory stimuli, performance did not significantly decline when paired with visual stimuli, but the pairing itself did not necessarily improve performance either. Thus, pairing visual and auditory stimuli together does not seem to be a useful strategy for improving recognition memory of individual items.

Additionally, pairing a stimulus in one modality with a familiar versus unfamiliar stimulus in another modality did not affect memory performance. Although the mere exposure theory suggests that associative memory should be better for familiar items over unfamiliar items (Zajonc, 1968), this did not seem to impact the memory of the individual items themselves. This does not necessarily imply that associative memory was unaffected by familiar versus unfamiliar pairings since our paradigm tested individual items instead of the item pairings themselves. However, the familiarity status of the paired item did not significantly impact item recognition.

While multimodal strategies are often recommended to cater to different learning styles (Felder & Silverman, 1988), the present study suggests differently. Based on these findings, we recommend that if you want to improve memory of individual items, you should *not* pair it with another stimulus. At best, there will be no effect, and it's possible that presenting items together actually disrupts the encoding process. Therefore, studying items without distractions seems to be the most effective strategy for enhancing item recognition. Future research can continue to investigate the nuances of sensory integration and its effects on different memory tasks to enhance performance in these practices.

References

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. Journal of Experimental Psychology: General, 142(4), 1323–1334. https://doi.org/10.1037/a0033872

Banks, W. P. (1970). Signal detection theory and human memory. Psychological bulletin, 74(2), 81.

Botta, F., Santangelo, V., Raffone, A., Sanabria, D., Lupiáñez, J., & Belardinelli, M. O. (2011). Multisensory integration affects visuo-spatial working memory. Journal of Experimental Psychology: Human Perception and Performance, 37(4), 1099.

Chang, S., Xu, J., Zheng, M., Keniston, L., Zhou, X., Zhang, J., & Yu, L. (2022). Integrating visual information into the auditory cortex promotes sound discrimination through choice-related multisensory integration. Journal of Neuroscience, 42(45), 8556-8568.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. Educational psychology review, 3, 149-210.

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. Proceedings of the National Academy of Sciences, 106(14), 6008-6010.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. Journal of Neuroscience, 22(13), 5749-5759.

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. Engineering education, 78, 674-681.

Gardner lab. (n.d.). https://gru.stanford.edu/doku.php/tutorials/sdt

Grill-Spector, K., Knouf, N. & Kanwisher, N. The fusiform face area subserves face perception, not generic within-category identification. Nat Neurosci 7, 555–562 (2004).

Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of Neuroscience, 17 (11), 4302–4311

Lloyd, Mary Anne; Appel, James B. PhD. Signal Detection Theory and the Psychophysics of Pain: An Introduction and Review. (1976). Psychosomatic Medicine 38(2), 79-94.

Macmillan, N. A., & Creelman, C. D. (2005). Detection theory: A user's guide (2nd ed. ed.). Lawrence Erlbaum Associates Publishers.

Morillon, B., Arnal, L. H., & Belin, P. (2022). The path of voices in our brain. PLoS Biology, 20(7), e3001742.

Ngo, C. T. (2013). Associative recognition memory and context effects using objects on natural backgrounds.

Nikel, L., Sliwinska, M. W., Kucuk, E., Ungerleider, L. G., & Pitcher, D. (2022). Measuring the response to visually presented faces in the human lateral prefrontal cortex. Cerebral Cortex Communications, 3(3), tgac036.

Paivio, A. (1969). Mental imagery in associative learning and memory. Psychological review, 76(3), 241.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. Trends in cognitive sciences, 12(11), 411-417.

Shams, L., Wozny, D. R., Kim, R., & Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. Frontiers in psychology, 2, 264.

Signal detection theory. (n.d.). https://www.cns.nyu.edu/~david/handouts/sdt/sdt.html

Swets, J. A., Tanner Jr, W. P., & Birdsall, T. G. (1961). Decision processes in perception. Psychological review, 68(5), 301.

Thesen, T., Vibell, J. F., Calvert, G. A., & Österbauer, R. A. (2004). Neuroimaging of multisensory processing in vision, audition, touch, and olfaction. Cognitive Processing, 5, 84-93.

Trébuchon, A., Alario, F. X., & Liégeois-Chauvel, C. (2021). Functional topography of auditory areas derived from the combination of electrophysiological recordings and cortical electrical stimulation. Frontiers in Human Neuroscience, 15, 702773.

Van Engen, K. J., Dey, A., Sommers, M. S., & Peelle, J. E. (2022). Audiovisual speech perception: Moving beyond McGurk. The journal of the acoustical society of america, 152(6), 3216-3225.

Weiner, K. S., & Zilles, K. (2016). The anatomical and functional specialization of the fusiform gyrus. Neuropsychologia, 83, 48-62.

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I. G., & Soltani, M. (1999). Recognition memory for faces: When familiarity supports associative recognition judgments. Psychonomic bulletin & review, 6(4), 654-661.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. Journal of personality and social psychology, 9(2p2), 1.

Appendix A: Familiar stimuli

Familiar face images:

1. Barack Obama
2. Donald Trump
3. Joe Biden
4. Michelle Obama
5. Justin Bieber
6. Dwayne Johnson
7. Kim Kardashian
8. George Washington
9. Morgan Freeman
10. Johnny Depp
11. Marilyn Monroe
12. Michael Jackson
13. Taylor Swift
14. Albert Einstein
15. Queen Elizabeth II
16. Beyonce
17. Rihanna
18. Britney Spears
19. Shakira
20. Elon Musk
21. Jackie Chan
22. Cristiano Ronaldo
23. Kylie Jenner
24. Lionel Messi
25. Selena Gomez
26. Ariana Grande
27. Bill Gates
28. Robert Downey Jr.
29. Mark Zuckerberg
30. John F. Kennedy
31. Lady Gaga
32. Leonardo Dicaprio
33. Elvis Presley
34. Steve Jobs
35. Princess Diana
36. Adolf Hitler

Familiar audios:

1. Happy Birthday
2. Fur Elise
3. Nationwide Insurance jingle
4. State Farm jingle
5. Twinkle Twinkle Little Star
6. Yankee Doodle
7. Pachelbel's Canon
8. Jingle Bells
9. Blue Danube Waltz
10. Eine Kleine Nachtmusik
11. Beethoven's Symphony No. 5
12. Despacito
13. Party in The U.S.A.
14. My Heart Will Go On
15. Hallelujah
16. Ave Maria
17. Bach Toccata and Fugue in D minor
18. William Tell Overture
19. Grieg's In The Hall of the Mountain King
20. Bizet Carmen
21. Eleanor Rigby
22. Bohemian Rhapsody
23. Gangnam Style
24. USA National Anthem
25. Alarm sound (Radar)
26. Alarm sound (Ripples)
27. Alarm sound (Slow rise)
28. Alarm sound (Old phone)
29. Super Mario Bros theme song
30. Scott Joplin's The Entertainer
31. Nokia ringtone
32. Netflix intro
33. Row Row Row Your Boat
34. Frere Jacques
35. The Wedding March
36. Swan Lake