wFebruary 28, 2022, V1

Visual Perception of 3D Space and Shape in Time

Part IV - 3D Shape Recognition by 3D Rotation

**Caominh Le, Samantha Pedersen, Nathaniel Chen, Jonathan Chan, Brian Ta,**

**Patrick Wilson, Trevor McCarthy, Emma Barseghyan, Anushka Chauhan, Hind Saif, Jonathan Tu, Darren Wijaya, Annika Zhang, Erica Li, Camille Marangi, Setayesh Nekarae, Felicia Wang, Alice Yanovsky, Umaima Afifa, Javier Carmona, Diego Espino,**

**Leonard Schummer, Philip Gudijanto, Gurleen Kaur, Andrew Lam, Matthew Mar,**

**Elizabeth Mills, Alexandra Nevins, Elijah Ortiz, Kyle Wheeler,**

**Aaron Blaisdell\*, and Katsushi Arisaka**

**New Authors: Daniel Beneda (Orcid: 0000-0001-9028-0832), Victor Li (0000-0001-8243-4264), Nyla Jafri (0000-0003-3549-826X), Yaochi Li (Orcid:0000-0002-2161-0364), Jackson Wissing (Orcid: 0000-0001-8294-8740)**

*University of California, Los Angeles*

*Department of Physics and Astronomy, Department of Electrical and Computer Engineering*

*\*Department of Psychology, UCLA Brain Research Institute*

*475 Portola Plaza, Los Angeles, CA 90095, USA*

*Contact:* [*arisaka@physics.ucla.edu*](mailto:arisaka@physics.ucla.edu)

# Abstract

Human vision has a remarkable ability to recognize complex 3D objects such as faces at any size, location, and orientation. If we initially memorize a face with a normalized size, front-facing, and at the object center, the direct comparison between the one-sized memory and an incoming new image would demand tremendous mental frame translations in 7D. How can we perform such a demanding task so promptly as we experience it in our daily lives?

This paper specifically addresses the recognition of human faces with arbitrary 3D orientation in the [Roll, Yaw, Pitch] axes. According to our new model of **NHT** (Neural Holography Tomography), space is represented by time utilizing the phase of an alpha brainwave. This principle should be applicable to any mental rotation in 3D; thus, it predicts that extra time is required to perceive a rotated face to revolve it back upright by the constant-speed alpha brainwave.

To assess this hypothesis, we designed a reaction time (RT) experiment, where participants were first asked to memorize sets of upright unfamiliar faces. Following the memorization phase, similar stimuli with a wide range of 3D rotation were presented, and RTs were recorded. As expected, the memorized upfront face had the fastest RT. The increase in RT was observed proportional to the rotating angle in all [Roll, Yaw, Pitch] axes. Roll had the flattest slope, whereas Pitch, especially in the upward direction, had the steepest. We propose that mental operations along the Roll axis require the least time because they can be conducted by linear frame translations on the log-polar retinotopy of the visual cortex.



# Introduction

Most people rely on vision as their primary source of sensory information for understanding and responding to their surroundings. Past research has mapped the complex network of brain structures that convey and process this information to at least thirty different cortical areas (Pessoa, 2014). Facial recognition, the ability to identify previously seen faces, is particularly key to facilitating everyday social interactions (S. Favelle & Palmisano, 2018). In fact, it is so fundamental that human infants are able to discriminate between familiar and unfamiliar faces at various angles (Turati et al., 2008).

Under the two-streams hypothesis, visual processing can be defined by two distinct pathways in the brain: dorsal and ventral (Goodale & Milner, 1992). Although recent evidence has shown that visual processing may not be as binary (Zachariou et al., 2014), within the two-streams hypothesis, the dorsal pathway constitutes the ‘where’ pathway of spatial perception and processes locational and orientational information, while the ventral pathway constitutes the ‘what’ pathway of object perception and facial recognition (Polanen, 2105). Moreover, the dorsal pathway has been shown to process information and guide actions without accompanying conscious knowledge, while the ventral pathway is involved in conscious perception (Fang & He, 2005). In the commonly cited ventral pathway model, visual information proceeds from the primary visual cortex (V1) through the anterior inferotemporal (aIT) cortex, where both the average receptive field size and latency increase (Kravitz et al., 2013). Cortical orientation vectors are only weakly selective in V1, where they receive input solely from the lateral geniculate nucleus (LGN) (McLaughlin et al., 2000), but they become more strongly orientationally selective throughout the pathway. In V3, a large majority of neurons are orientationally selective (Felleman & Van Essen, 1987).

It has been long supported that the ventral pathway processes semantic information by a log-polar coordinate system in humans and primates (Arcaro et al., 2009; Engel et al., 1997; Kolster et al., 2014). However, the adaptive benefit for this distortion of the visual cortex has not been fully understood in the context of object recognition, especially with the failure of the current bottom-up theory of visual processing to address the efficiency and accuracy of human facial recognition. A top-down signal connecting visual input and internal imagery is necessary to allow rapid recognition of familiar faces at any size, location, or rotation in as little as 200 ms (Caharel et al., 2015; Dijkstra et al., 2017). Formed from contextual reinforcement, top-down pattern recognition also allows for efficient analysis and categorization of key facial features (Puce et al., 1999).

The viewpoint dependence of facial recognition on Roll, Yaw, and Pitch has been previously studied, and performance has been found to be most optimal in the Roll axis, followed by Yaw and Pitch (S. K. Favelle et al., 2011). Another study found that participants assigned to a front or ¾ yaw comparison view performed better than those assigned to pitch-up or pitch-down views, possibly due to how these angles maximize the contour and projection of feature information, allowing better extraction of 3D information (S. Favelle & Palmisano, 2018). In general, faces are processed holistically, but when skewed at different angles, the visual processing mechanism of the cortex is forced to focus on individual key features for recognition (S. Favelle & Palmisano, 2012). Ultimately, holistic processing requires significantly more time than digesting a few key features but may lead to better accuracy.

We have recently proposed the new concepts of MePMoS (Memory-Prediction-Motion-Sensing) and NHT (Neural Holographic Tomography) to map the processes of this conscious top-down approach (Arisaka, 2022a, 2022b; Arisaka & Blaisdell, 2022). Facial recognition occurs when a relevant memorized pattern of the stimulus undergoes 7° of transformation to overlap with the incoming image, and this transformation can be expressed as measurements in the latency of alpha brainwaves. In the past, alpha oscillations have been studied as a possible mechanism to encode visual input by inhibiting irrelevant brain structures and directing information to appropriate neural structures for semantic encoding during early perception or semantic retrieval during object recognition (Klimesch et al., 2011). Under the MePMoS model, we predict that alpha brainwaves are also involved in scaling and rotating memorized data through variable phase shifts determined by a time delay to overlap incoming signals for recognition. After the initial visual input is received during an alpha brainwave cycle, the relevant memory is extracted and transformed by alpha brainwave phase shifts to predict the appearance of the incoming stimulus. Upon overlap, two more alpha cycles are required for conscious awareness of the congruency and instructions for motion output, placing the simplest conscious recognition and decision-making at around 400 ms.

The NHT model also effectively explains the significance of the log-polar distortion of the ventral pathway. Conversion of incoming Cartesian images to log-polar makes the projection onto the V1 visual cortex scale and roll-rotation invariant. Rescaling the memorized image, which causes key features to move to or from the periphery, becomes a simple horizontal translation of the distorted projection on the log axis. Likewise, rotating the stimulus along the Roll axis becomes a simple vertical translation on the polar coordinate axis. These translations can be encoded through phase shifts of top-down signals on the timescale of constant-frequency alpha brainwaves. Larger rotations about the roll axis require larger translations in the projected V1 image, leading to higher phase shifts in alpha brainwaves and time delays on the order of 100 ms.

Unlike Roll, Yaw and Pitch involve more extensive stepwise transformations to overlap memorized data to incoming stimuli. The brain must first translate the memory from an egocentric frame of reference to an object-centric frame. Once centered on the object, the memory can be rotated along the Yaw or Pitch axes. Finally, the frame can be restored to the original egocentric frame of reference and compared to incoming stimuli from the same frame. While all three steps may require shifts in brainwave patterns, only the rotation step differs between stimuli, given that they were memorized at the same distance and location. This transformation can be observed by reaction time differences, with larger rotations leading to higher alpha brainwave phase shifts and latencies about 100 ms apart. Furthermore, increasing familiarity with an object should decrease intrinsic reaction times, regardless of the cortical pattern, because the speed of mental frame transformations remains constantly dependent on the speed of brainwaves.

# Results

Reaction times were recorded for rotation of four categories of stimuli: unfamiliar faces, familiar faces, characters, and words. Data was analyzed separately for negative angles that consisted of downwards Pitch, leftwards Yaw, and counterclockwise Roll and their positive counterparts. 0° and 180° (for Roll) were analyzed in both data halves.

Data at each rotation angle was averaged separately for individual participants, and linear fits were calculated for negative and positive angles using minimum chi-square estimation. Gamma distributions were generated from histograms of the reduced chi-square values (Supplemental Information). Gamma distribution modes near 1.00 suggested strong linear fits for most individual participants, justifying linear analysis for the aggregate data pool (Table 1).

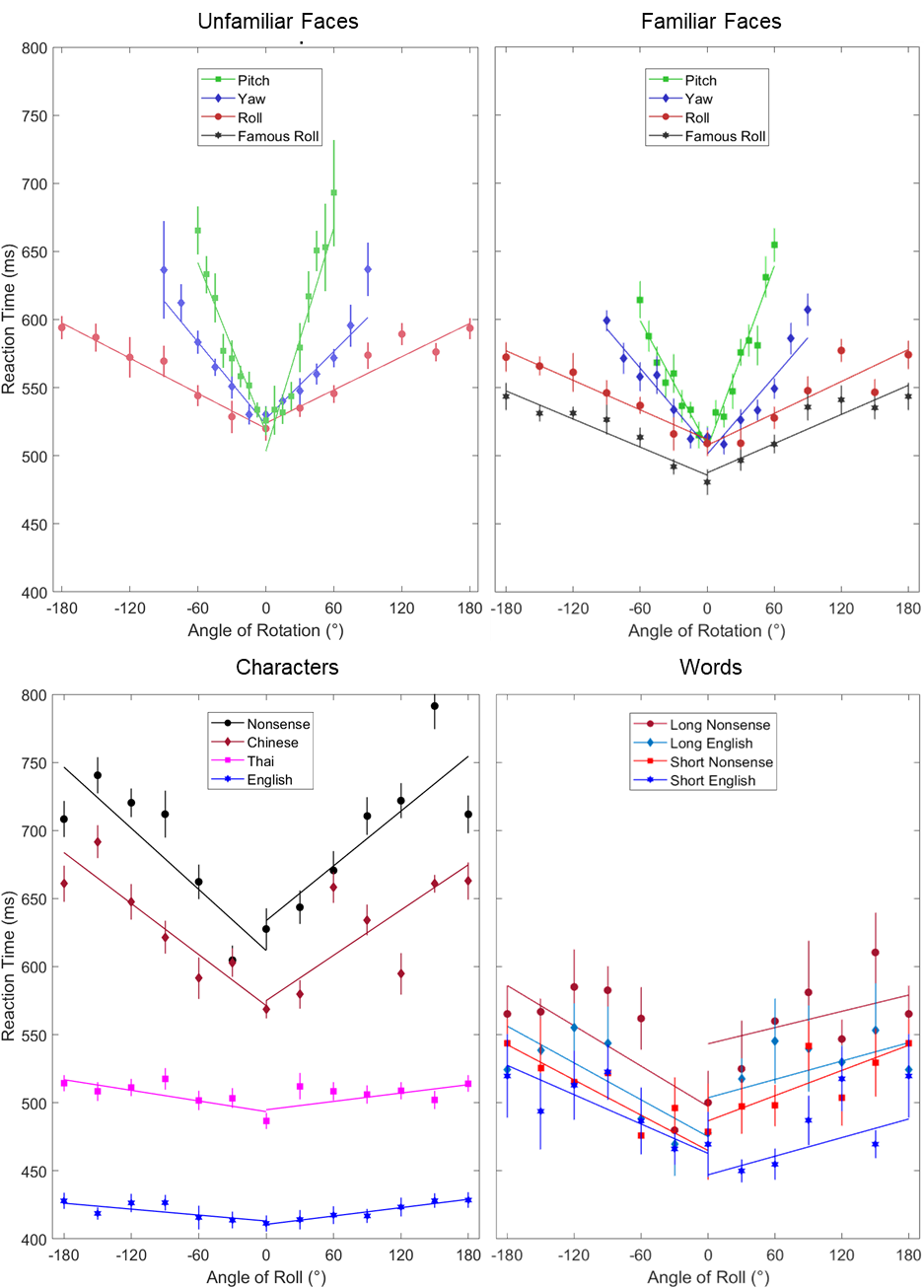
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Protocol | Group | n | Individual Reduced χ2 Modes | Protocol | Group | n | Individual Reduced χ2 Modes |
| Face Roll | A | 14 | 1.04 | English Roll | A | 14 | 1.15, 3\* (bimodal) |
| B | 15 | 1.18 | B | 15 | 1.02 |
| C | 13 | 0.94 | C | 13 | - |
| D | 14 | 1.31 | D | 14 | 0.89 |
| Face Yaw | A | 14 | 0.47 | Thai Roll | A | 14 | 2.7\* |
| B | 15 | 1.04 | B | 15 | 0.70, 2.8\* |
| C | 13 | 1.17 | C | 13 | - |
| D | 14 | 0.96 | D | 14 | 1.33 |
| Face Pitch | A | 12 | 1.66 | Chinese Roll | A | 12 | - |
| B | 15 | 1.76 | B | 15 | 2\* |
| C | 13 | 1.23 | C | 13 | - |
| D | 14 | 0.67 | D | 14 | 1.11 |

**Table 2** Gamma Distribution modes for Reduced Chi-squares of linear fits to individual participant data. (\*) Some distributions were more uniformly distributed, skewed, or had an additional mode. For group A, English Roll was bimodal, and Thai Roll was skewed, giving it a smaller calculated mode of 1.33 than the estimated actual mode. For group B, Thai Roll was bimodal while Chinese Roll had a lower calculated mode of 1.30 than the estimated actual mode.

Individual mean reaction times were averaged from the aggregate data pool. Standard errors were calculated after normalizing individual participants’ data around the overall average by Cousineau normalization (Morey, 2008). Linear fits were calculated for negative and positive angles using minimum chi-square estimation. Reduced chi-squares greater than one suggested the possibility of more optimal non-linear fits while models with reduced chi-square less than one were judged based on the size of error bars. Plots showing stacked individual participant data before and after normalization can be found in the Supplementary Figures along with reduced chi-square histograms. An example of data analysis plots is shown in Figure 2.

## Faces at a Glance

Reaction times were obtained for both unfamiliar and familiar faces. The unfamiliar faces group consisted of four subgroups who took separate protocols (Supplemental Analysis A). Because of consistencies between results for each group, the data was pooled together as a single aggregate group (Figure 1). Differences and consistencies between groups are considered in Supplemental Information. Familiar faces were tested by training participants to recognize unfamiliar faces or by testing participants with famous faces already known to them, which included US presidents and famous celebrities. Prior to the experiment, participants confirmed that they knew the identity of the famous figures.



**Figure 1** Comparison of Unfamiliar and Familiar protocols. Among the faces, Roll was the most symmetric and least steep while Pitch was the most asymmetric and the steepest. There was a significant decrease in reaction time for famous faces compared to unfamiliar faces and a smaller decrease when comparing unfamiliar and familiarized faces.

### Unfamiliar Faces results in Linear V-Shape Curves

All three unfamiliar face protocols resulted in similar average y-intercepts and V-shapes with Roll being the least steep and Pitch being the steepest. Roll had a symmetric V-shape while Yaw had a slightly asymmetric V-shape and Pitch had a very asymmetric V-shape. All three protocols had small reduced-chi-square values for negative angles, indicating strong linear fits. Interestingly, the reduced chi-square values for positive angles were greater, indicating weaker but acceptable linear fits. Both roll and yaw deviated from the linear trendline at the higher angles with a slight decrease in reaction time around ±150° roll and an upward curve towards ±90° yaw. In the pitch protocol, there was a downward deviation from the linear prediction at 22.5°.

### Familiar Faces Rotation is parallel to Unfamiliar Faces

Like unfamiliar faces, consistent average y-intercepts and V-shapes were found for the trained face protocol. Again, Roll was the least steep and Pitch was the steepest. There was a decrease in intercept of approximately 25 ms from the unfamiliar face protocols. The decrease in intercept was more significant between the unfamiliar and famous faces, with a 50 ms decrease in intercept. Nevertheless, roll protocols remained parallel between familiarity levels as well as the yaw protocols. Upward deviations from the linear fit at larger angles were less pronounced for trained yaw, but trained roll still showed a decrease in reaction time at 150°. Trained pitch was less steep than unfamiliar pitch, with a decrease in 0.5 ms/°.

## Language Recognition

The same reaction time experiments were conducted for characters and words. Four types of characters were tested by combining differing levels of familiarity (english vs. foreign) and complexity (simple strokes vs. superimposed strokes). Words recognition was tested similarly with a matrix of english vs. nonsense words and short vs. long words. All participants confirmed that their primary language was English prior to the experiment.

### Simple Characters are Flat, Complex Characters are like Faces

Both English and Thai roll protocols were symmetric and almost flat, being 4 times less steep than faces roll. Thai was roughly parallel to English but took on average 85 - 90 ms longer at any given angle and had an intercept between trained and familiarized face rotation — participants responded 30 ms faster to unrotated Thai than unrotated unfamiliar faces. Both linear predictions were strong fits as indicated by small reduced chi-squares, especially for negative angles. As with faces roll, there was a drop in reaction time around ±150° for Thai. Chinese and Nonsense roll were also symmetric and linear, but they had slightly steeper slopes than face roll, with a 0.25 ms/° difference. Intercepts were also larger by 50 ms for Chinese and 100 ms for Nonsense characters. Data at 120° for Chinese and 150° and 180° for nonsense seemed to deviate from the linear trend as possible outliers.

### Words rotation is comparable to Faces Rotation

Shorter and sense word protocols had lower intercepts than longer and nonsense word protocols, which were comparable to the intercept seen for faces rotation. Nevertheless, all four word experiments resulted in similar slopes to faces roll. Interestingly, they were slightly asymmetric, especially for the complex longer english and nonsense words, with counterclockwise rotation being 0.2 – 0.25 ms/° steeper than clockwise rotation. However, words rotation data remains preliminary with a small sample size. While reduced chi-square values were small, error bars were larger, indicating greater deviation between the limited number of participants.

## 2.3 Summary Comparison

### Comparison of Protocol Intercepts

Within Groups B, C, and D, intercepts remained mostly constant between protocols, but Group C intercepts of around 540 ms were notably higher than Groups B and D around 510 ms. Group A intercepts were also comparable to group C except for face pitch, which was faster than all other groups. Among character protocols, English had the fastest intercept time at 420 ms; Thai had an intermediate intercept time at 500 ms; and Chinese had the slowest intercept time at 575 ms.

## Familiarity Affects Only Intrinsic Recall Time

The effect of familiarity on face recognition was tested using the same gamified protocol used for Group D participants with familiar and famous faces. Familiarity decreased reaction time independently of the degree of rotation in all three axes, especially for famous faces. The intercepts for all axes decreased between 10 and 20 ms when the faces were familiarized through training. In contrast, participants responded faster to famous faces by approximately 35 ms. Nevertheless, slopes remained parallel between unfamiliar, familiar, and famous faces.

### Comparison of Protocol Slopes

Slopes for Face Roll and Yaw between groups were consistently around 0.4 and 0.9 ms/°, respectively. On average, the V-shape for Roll was symmetric while for yaw, it was slightly asymmetric with leftward yaw angles leading to longer reaction time shifts. Additionally, yaw was twice as steep as Roll. In contrast, face pitch was consistently asymmetric between all groups, with upward rotations leading to longer reaction time shifts than downward rotations.

Table

Description automatically generated with medium confidence

**Figure 7** Comparison of the influence of rotation on reaction times between unfamiliar, familiar, and famous faces for each protocol, represented by trendline slopes.

A picture containing table

Description automatically generated

**Figure 8** Comparison of reaction times for unrotated stimuli between unfamiliar, familiar, and famous faces for each protocol, represented by trendline intercepts.

# Discussion

## Linearity of Face Recognition in Three Rotational Axes

To support the new model of MePMoS and NHT, reaction time experiments were conducted for unfamiliar faces and character rotation along the Roll, Yaw, and Pitch axes. Four independent groups, including two internal and two external groups, took data with variations between each group's protocol and experimental environment. Nevertheless, agreements were found between corresponding protocols from each group. A positive linear correlation was found between reaction time and the degree of rotation of unfamiliar faces in all three axes, with the memorized front-facing, level, and upright faces having the fastest reaction time around 525 ms. Larger angles of rotation also have longer reaction times on average. These results support the NHT model, providing a way for the brain to transform memories of the faces by alpha brainwave phase shifts to overlap with incoming stimuli through time delays. Larger phase shifts caused by longer time delays could allow for recognition of larger rotation of the face in all three axes. The timescale of reaction times on the order of 100 ms also support the involvement of alpha brainwaves, with a 525 ms shift for unrotated stimuli, a 75 ms shift for maximal roll, a 100 ms shift for maximal yaw, and a 125 ms shift for maximal pitch.

As in previous studies, reaction times to rotations along the roll axis were generally faster than yaw and pitch (S. K. Favelle et al., 2011). The conversion to polar in the visual cortex V1 could aid in hastening roll rotation, making Roll faster than the other two transformations. Roll rotation also occurs in 2D, so the log-polar ventral pathway is primarily needed to perform the rotation. In contrast, rotation along the pitch and yaw axes requires the brain to rotate a 3D memory of the stimulus, which needs input from both the dorsal and ventral pathways (Todd, 2004). With involvement from both pathways, frame transformation between egocentric and object-centric references before application of the Yaw or Pitch rotation may contribute to the steeper reaction time shifts. Between Yaw and Pitch, Yaw was generally faster as previously observed (S. K. Favelle et al., 2007). By repeated practice or evolution, humans may have adapted to recognize yaw rotations more efficiently, likely due to the tendency to interact with others at a level height, leading to less exposure to faces at extreme pitch orientations than yaw orientations.

## Unfamiliar vs Familiar Stimuli

As hypothesized, familiarity had a decreasing effect on reaction time for faces, causing a decrease in latency across all degrees of rotation between each axis. Training resulted in 10-15 ms time decreases while innately familiar famous faces were 35 ms faster on only the roll axis. Familiarity appeared to affect all three axes of rotation to the same degree, revealing the possibility of discrete decreases in reaction time resulting from bypassing pathways in decision-making. This could occur in the basal ganglia, which is a major control center for motor movement. Increasing the linkage between different facial features could expedite the recognition process by a discrete 15 ms, which may be further increased by an additional 15 ms for famous faces.

The faces used for famous Roll were different from those used in the unfamiliar and familiarized protocols. Hence, the decreased reaction time may have arisen from greater ease in distinguishing between the three faces. However, given that training had an effect on reaction time, it is likely that familiarity contributed to this effect. Future experiments should be completed to further understand this effect.

Face recognition could be optimized by scanning in a triangular pattern from the eyes, nose, and mouth. In contrast , complex characters were more difficult to recognize due to a lack of recognizable key features, resulting in a greater central intercept for chinese and nonsense characters. Nevertheless, the same brainwave frequency is likely used to scan incoming visual input due to the parallel slope with faces.

Similarly, the nonsense characters, which were created by scrambling simple english words, had significantly higher intercepts and slopes than the short english words. The 2D nature of the characters and lack of systematic scanning in a single direction as in words could have contributed to these increased parameters. One possible mechanism by which rotated words were processed faster could be parallel processing of letters in the words due to their 1D arrangement, allowing the brain to focus on the predicted location of subsequent characters. In result, word processing was more like facial rotation in that it consists of key features in predictable locations. Furthermore, the recognition time was indirectly positively correlated with word length, which could increase the number of key features that required parallel processing.

As complexity and familiarity were shown to have different effects on recognition time, there likely exists separate mechanisms by which they can affect visual processing. Complexity likely affected the frequency by which stimuli were processed, given the significant difference between the slope of simple characters and complex characters. Simple stimuli, like english and thai characters or key parts of faces such as the eyes, nose, and mouth, could be processed locally and in parallel by high-frequency gamma waves at 100 hz. More complex stimuli, like entire faces, words, or complex characters with several key features, may be processed by slower beta or alpha brainwaves at 20 or 10 hz.

Both complexity and familiarity affected intercept reaction time, which could depend on processing within the basal ganglia, a major decision-making center of the brain that influences movement. With its layered composition, each layer of the basal ganglia could act as a filter to categorize memories with increasing detail and specificity for memories within deeper regions of the ganglia. Hence, as with words, more complex stimuli with a greater number of key features and details may be specified in the deeper layers while less detailed and simpler stimuli may be more accessible at upper layers.

## Systematic Error and Non-linear and Asymmetric Trends for Faces

Roll rotations above 120° were nonlinear. While there is a possibility that at such angles, the faces were still processed holistically yet very insufficiently (Richler et al., 2011), this finding is better explained by a qualitative shift in performance stemming from the loss of coherence between different facial features, which causes facial recognition to depend on some other non-holistic mode of processing. Upon reaching larger rotations, faces may be processed as objects, which suggests that faces may be processed less as a whole but by faster recognition of a few key features (Rossion, 2009). As the most motivated experimental group, Group D participants may have adopted this more “active” recognition to score more points when “passive” holistic processing took too long. Nevertheless, in daily life, it is likely that “passive” holistic processing through phase shifts of alpha brainwaves is more prevalent due to its higher accuracy.

Pitch had a significantly decreased reaction time at 22.5°. This could result from different heights of participants to which the headrests were adjusted to allow comfortable data-taking. On average, participants may have been slightly above eye level with the center of the screen, leading to the central reaction time when the stimuli faces were looking slightly upwards. However, given the distance between participants and the screen, this would only contribute to a small shift in the central angle. A more plausible cause could be the difference in overhead and ambient lighting between the professional and remote data-taking, which could influence one's perception of the faces (Favelle et al., 2017).

Yaw was slightly asymmetric while Pitch was highly asymmetric. For Yaw, this could represent systematic error from imperfectly symmetrical database faces, resulting in easier feature detection on one side of the face compared to the other. Likewise, differential lighting could also lead to differences in left and right face perception. On the contrary, asymmetry in Pitch can be explained by more exposure and practice in recognizing and processing more features from the top of other people’s heads than their chins.

## Further Studies

In the current experiment, participants reacted fastest to the unrotated faces that they memorized. However, it is not known whether the brain automatically stores memorized faces at 0° or at the most practiced rotation. If the latter holds true, the observed V-pattern may translate horizontally to the angle at which training occurred. On the other hand, it is possible that the brain is biologically hardwired to automatically memorize novel faces as upright, front-facing, and level to facilitate daily social interactions.

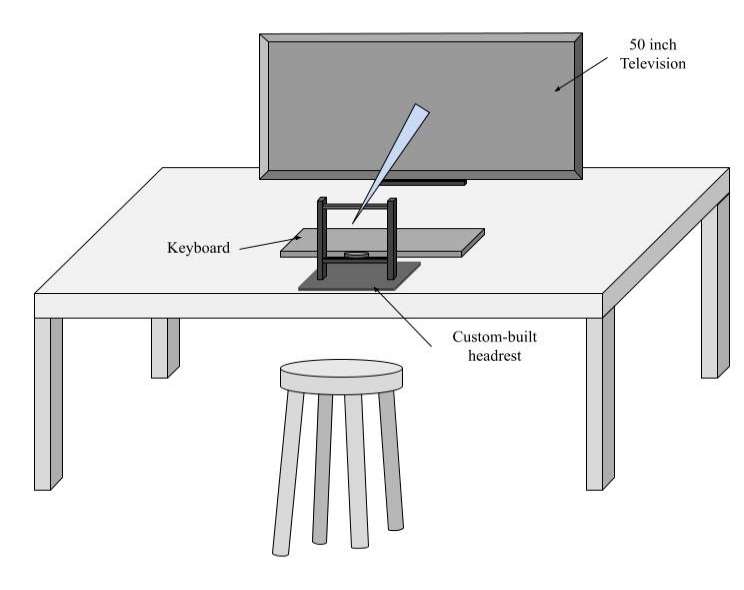
Future experiments should also combine multiple transformations, such as multiple rotational axes of rotation and scaling. The transformations may simply add linearly or have cooperative or opposing effects, especially Yaw and Pitch rotations, which require similar steps to perform. The transformations may also contribute asymmetrically or even be independent where the final reaction time is determined by the alpha phase shift that normally takes more time. Finally, reaction time experiments could be used to better understand the encoding of transformation of alphabets and words, which may yield similar results to faces. This could be due to the transformations being the combination of multiple familiar Gamma-encoded characters, paralleling facial features in facial recognition.

This research supports a novel model of top-down human recognition of unfamiliar faces and the nuanced neurological pathway it adopts. We have provided evidence that visual processing involves the conversion of 7° of transformation of images to time differences via a top-down pathway and the involvement of alpha and gamma brainwaves phase shifts in encoding rotation. Top-down processing may better explain the incredible efficiency of stimulus recognition over bottom-up processing and how humans evolved to become intelligent beings that could recognize other individuals and comprehend language.

# Materials and Methods

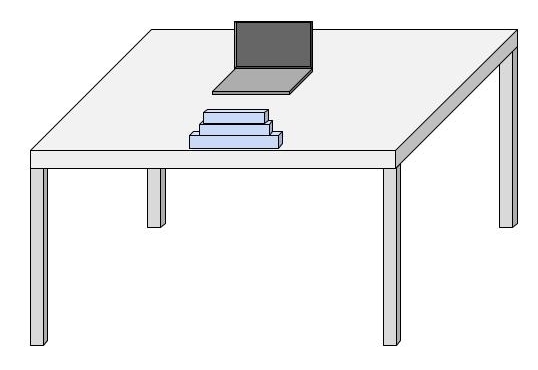
## Participants and Equipment

Four groups of undergraduate students at the University of California, Los Angeles (UCLA), all of which had a normal or corrected-to-normal vision, participated in the data-taking process for unfamiliar faces and characters. Due to the difficulties of recruiting participants during the COVID-19 pandemic, lab members from the UCLA Elegant Mind Club who designed the experiments also took data as internal participants. Groups A (n =13) and B (n = 15) consisted of internal participants involved with designing the experiment, respectively, who took data using personal computers and monitors of various sizes (Figure 8 Left). Participants created headrests from books or other household objects. Group C (n = 13) and D (n = 14) consisted of external participants, respectively, who took data in a professionally controlled laboratory setting using a 50-inch television, Sony headphones, and a custom-built headrest (Figure 8 Right). Participants for Group C were members of the Elegant Mind Club but were uninformed of the experiment and hypotheses while participants for Group D were completely external and recruited from lower division Physics lab classes. These groups’ data were assessed separately for consistency, then combined in a combined analysis as shown in the Results. Approval was given by the UCLA Institutional Review Board (IRB#19-001472-AM-00003) to conduct shorter experiments with external participants. All participants provided consent before completing the study and were debriefed afterward. A new group of internal participants, Group E, took more data on characters, familiar faces, and words.



**140 cm to screen**

**Headphones**



**Homemade Headrest**

**140 cm to screen**

**Sound from speakers or headphones**

**Laptop/monitor screen size varied**

**Figure 8** Drawing of experimental setups. (Left) Remote participants used makeshift headrests from household items to position their heads 50 cm from their laptops or monitors with a keyboard within arm's reach. (Right) Local participants took data in a controlled setting with a custom headrest 140 cm from a 50'' TV and a keyboard within arm's reach.

## Display Calibration and Reaction-Time Consistency

To ensure consistency across devices, remote participants ran a size and time calibration script to control figure display size and to account for monitor-delay parameters for the subsequent protocols. The size calibration determined the screen's length to pixel conversion factor by prompting participants to resize vertical bars, letters, and images to various sizes using a ruler. The time calibration determined the input lag by showing a Gabor pattern on screen through PsychoPy and prompting users to press a key while recording themselves with a high-speed camera. A plot of PsychoPy times vs. actual times from the recording was generated to determine the time delay at the y-intercept.

## Stimuli Selection

### Unfamiliar Faces

Each face rotation experiment involved 9 emotionally-neutral adult face models, all of which were Caucasian and female, that were selected from the Stirling/ESRC 3D Face Database and exclusively grouped into sets of three based on similar age, face texture, skin tone, and head shape (*Stirling ESRC 3D Face Database*, 2021). Faces with less readily identifiable skin texture (freckles, moles, wrinkles) were preferred for experiments. Images were converted into grayscale to minimize the effects of differences in skin tone and blemishes shown on a gray background. The models were rendered in Blender, an open-source 3D graphics software, with flat lighting and Gouraud Shading. Renders of the roll protocol were captured in 30-degree intervals from -150 to 180°; the yaw protocol was captured in 15-degree intervals from -90 to 90° and the pitch protocol was captured in 7.5-degree intervals from -60 to 60°.

### Repeating faces and Randomization

For group A, the roll, yaw, and pitch protocols each had a different set of faces. For groups B-D, the protocols featured the same three sets of faces to minimize the systematic effect of using different faces. To negate the effects of the participants learning and familiarizing themselves with the faces throughout the three protocols, participants were randomly assigned one of six permutations to run the face protocols. Stimuli images can be found in the repository linked in the Notes about Data and Authors section.

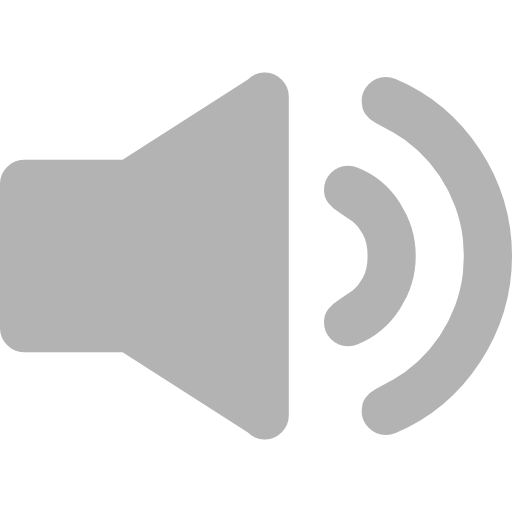
### Familiarized and Famous Faces

Faces were familiarized through a structured, but unsupervised, 3-day training period where participants were tasked with memorizing the faces shown at different angles of roll, yaw, and pitch for 30 minutes per day. Fun facts and nicknames were presented alongside the faces to assist with memorization.

This study also examined the effect of roll rotation for even more familiar famous faces, using U.S. presidents and famous celebrities. Like unfamiliar and familiar faces, all faces appeared white on a gray background in 30-degree intervals of roll rotation from -150 to 180°.

## Experimental Procedures

All experimental procedures followed the same general approach. In the learning period, participants were shown the unrotated stimuli for a duration of time after visual instructions to press a specific key whenever the stimulus was shown on screen. In the practice period, participants were assessed with unrotated stimuli without recording accuracy or reaction times; while in the testing period, participants were tested with rotated stimuli, and accuracy and reaction times were recorded. Each trial began with a white cross flash at the center of the screen followed by a random pre-trial delay seconds before the stimulus appeared at the center of the screen (Figure 9). After pressing a key, the participants received feedback on whether they answered correctly or incorrectly. If participants responded incorrectly during testing periods, the missed trial was repeated later at the end of the set so that all angles were evaluated equally. This routine was repeated for each of the three sets of three stimuli with a 20-second break between each set. All faces and characters were displayed at an 8-degree pitch eccentricity from the eye’s opening angle. Participants from groups C and D were shown a tutorial video before each data-taking session. Differences between each group are explained below and summarized in Table 4.



**Feedback**

**Participant Response**

**500 ms Interstimulus Period**

**Variable Pre-trial Delay (500 - 1500 ms)**

**Cross Flash (200 ms)**

**Post-trial Feedback**

**Figure 9** Flowchart showing steps for each trial. All faces and characters trials followed the same general steps, but there were intergroup differences in how participants responded and the audio/visual post-trial feedback

### Remote Unfamiliar Face Data during COVID-19

In the learning period, only the attractor face or character was shown on screen for 30 seconds and mapped to the letter 'v'. Participants were instructed to press 'n' when the unshown distractor faces and characters appeared. The practice periods consisted of 35 randomized trials with an equal probability of showing attractor and distractor stimuli, meaning attractors appeared twice as often as each individual distractor, and the testing periods consisted of 50 randomized trials. Halfway through the testing period, participants were given a shorter practice period with 15 trials. A solid A4 tone played for 0.5 seconds when participants pressed the wrong button. No sound played for a correct response.

### Group B: Full 3-Choice Reaction Test

In the learning period, all three faces or characters were mapped to corresponding keys (v, b, or n). Participants were instructed to press a key before and after the corresponding stimulus appeared for 10 seconds. The learning period was repeated for a total of 20 seconds for each face. The practice periods consisted of 12 randomized trials with an equal probability of each stimulus appearing. For the full-length procedure, the testing periods consisted of 50 randomized trials with a shorter 6-trial practice period at the halfway mark. For the shortened procedure, training and practice periods were identical, but the testing periods consisted of 32 trials without an interim practice period. The same audio feedback was used as in the attractor-distractor test.

### Group C: Shortened 3-Choice Reaction Test

Training and practice periods were identical to the full 3-choice reaction test. The testing periods were shortened to 32 trials per set without an interim practice period. Additionally, all three protocols were taken within one hour in a preset random order with a 1–4-minute break between each protocol. The audio feedback was used in the attractor-distractor test.

### Group D: Shortened 3-Choice Reaction Game

Gaming features were added to the training, practice, and testing periods of the shortened 3-choice reaction test. The instructions were modified to be more game-like with textual changes and animations a short demonstration of the current rotation being assessed, a scoreboard showing 10 high scores, and a level screen. The scoring system was created with correct responses earning between 400 and 1000 points depending on response speed. A green message showed the score increase and a major triad played beginning with an A4 root that increased with consecutive correct answers. Incorrect responses resulted in a loss of up to 400 points with faster responses having a smaller decrease. After 1.2 seconds, the trial would time out, resulting in an automatic 400 score decrease. A red message showed the score decrease and an A4 tone followed by an F4 provided a defeat sound effect.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Group Letter | Source of Participants | Max n\* | Equipment | Experimental Design |
| A | Internal members from Club | 14 | Laptops and various-sized monitors | Attractor vs. Distractor Test |
| B | Internal members from Club | 15 | Laptops and various-sized monitors | Full 3-choice Reaction Test |
| C | Semi-External members from Club | 13 | 50'' TV, headphones, and headrest | Shortened 3-choice Reaction Test |
| D | External participants from Physics labs | 14 | 50'' TV, headphones, and headrest | Shortened 3-choice Reaction Game |
| E | Internal members from Club | 22 | 50'' TV, headphones, and headrest | Shortened 3-choice Reaction Game |

**Table 4** Descriptions of four data taking samples. (\*) As protocols were completed separately in a less controlled fashion, some internal participants did not complete all protocols. External participants also completed protocols over two days, so some only completed half of the protocols.

## Statistical Analysis

Reaction time trial data were categorized by their stimulus heights, and further analysis was done within the data separated by each height (Ta 2022; Bustanoby 2022; Afifa 2022).

### Error Analysis

The resulting reaction times were averaged, and the error bars were plotted using the calculated standard error: taking the standard deviation of the eccentricity angle reaction time divided by the square root of the number of trials within the eccentricity angles. After reaction time data were categorized by angle and outliers were removed, the standard deviation and standard error were calculated, with the standard error being used in the plots.

### Outlier Analysis

Outlier data were removed before averaging the RT at each eccentricity. Data points that fell below the 25th percentile minus 1.5 times the interquartile range or above the 75th percentile plus 1.5 times the interquartile range for reaction time data at each angle were removed. The resulting data points were then averaged and used in chi-square calculations.

75thpercentile + 1.5 ∗ IQR < points excluded < 25thpercentile − 1.5 ∗ IQR

### Hypothesis Testing

To ensure that an attention shift occurred during each choice RT block, response accuracy was analyzed using a one sample t-test for a mean of one, with three being the number of available choices for each protocol. The test was performed for each stimulus height and a participant’s data was removed for a given height if the corresponding p-value was greater than 0.05.

### Normalization of average RT

To better represent the relationship between RT and stimulus height in the aggregated analysis, we normalized each participant’s data to a global average for each protocol (Morey). We normalize the data by subtracting the appropriate participants' mean performance from each observation and then adding the overall mean score to every observation. For example, if yij is the i-th participant’s average reaction time in the j-th condition then the normalized observation, zij, can be defined as:

### Line of best fit and correlation coefficient (χ2)

A best-fit line was added using a least-squares linear regression for both positive and negative heights and their corresponding reaction times, with the memorized reference size being included in both positive and negative fits. The linear regression returned slope and intercept parameters as well as the correlation coefficient and a two-sided p-value in which the null hypothesis assumes a slope of zero. Additionally, the standard error of the fit was returned.

A best-fit line of the form y = mx + b was calculated using chi-square minimization to find the best-fit slope and intercept parameters for both positive and negative heights with each side including the reference height. From the chi-square minimization, error estimates for the slope and intercept parameters were computed by holding the other parameter constant at the best fit and calculating which parameter values would yield the minimum chi-square value + 1. This process was done for both slope and intercept parameters. The reduced chi-square was also calculated to determine the goodness of fit by dividing the minimum chi-square value by the° of freedom. The correlation coefficient was also calculated using the NumPy library in Python.

where is defined as the mth residual and is the standard error of the reaction time data for an angle m.

where *ν* is the° of freedom.

# Notes about Data and Authors

**Data availability**

The experimental data, stimuli, and code to calibrate the monitors/TVs, analyze data, and run the experiments can be found at [https://github.com/UCLA-Elegant-Mind-Club/EMC-Rotation-Station](https://github.com/cm600286/EMC-Rotation-Station) and <https://github.com/UCLA-Elegant-Mind-Club/EMC-Scrotation-Station>.

**Competing interests**

The authors declare no competing interests.

**Acknowledgments**

This work was in part supported by the Dean’s Office of Life Science, Dean’s Office of Physical Science, Chair’s Office of the Department of Physics and Astronomy, and the Instructional Improvement grant by the Center for the Advancement of Teaching, all of which are at the University of California, Los Angeles. We would also like to extend our gratitude to the Arisaka Lab and the UCLA Physics & Astronomy for providing us with working space and other resources. We are grateful for all those who have provided us with human face image/model repositories. Finally, we thank the banana for being a good model of holography.

**Author contributions**

KA developed the concept and supervised the project. C.L., S.P., C.M., B.T., H.S., E.B., S.N, J.C., J.T, A.C., A.Z., D.W., N.C., E.L., G.K., E.O., F.W., A.Y., A.N., K.W., and P.W. wrote the manuscript. The research was conducted by three groups. In the "senior" group, N.C., B.T., E.M., E.L., S.N., F.W., A.Y., P.W., T.M., U.A., C.M., G.K., A.N., E.O., K.W., M.M., A.L., and L.S. designed and conducted initial experiments for experimental Group A. The "new" group, comprised of C.L., N.C., J.C., S.P., B.T., D.W., A.Z., J.T., A.C., E.B., and H.S., redesigned the experiments with the help of B.T and P.G. and conducted experiments as operators for groups B, C, and D. The “familiar” group, D.B., V.L., N.J., J.W., and Y.L., redesigned the experiments to include varying degrees of familiarity with the stimulus images. D.E. managed IRB approval and participant recruitment, and D.E., J.C., and new group members set up hardware for external data taking. B.T, P.W., T.M., and C.L. created PsychoPy and MATLAB scripts. P.G. and N.C. created stimulus images for the senior group, and N.C., C.L., J.C. and A.Z., selected and created stimuli for the new group. N.C., B.T., E.L., S.N., E.O., U.A., and P.W. performed statistical analysis for the senior group, and C.L., S.P., J.T., A.C., and D.W. performed final statistical analysis and figure generation. N.C., E.L., S.N., F.W., and A.Y. coordinated the senior group process, and C.L., J.C., S.P., and D.W. coordinated the new group progress. K.A. and B.T. supervised the work, and K.A. and A.B. finalized the manuscript. All authors have looked at and approved the manuscript.

# References

Afifa, U., Baik, J., Biju, A., Carmona, J., & Chung, S. (2022). *Visual Perception of 3D Space and Shape by 7D Frame Translation in Time Part I: Two-dimensional Space Perception by Two-dimensional Linear Translation*.

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic Organization of Human Ventral Visual Cortex. *Journal of Neuroscience*, *29*(34), 10638–10652. https://doi.org/10.1523/JNEUROSCI.2807-09.2009

Arisaka, K. (2022a). *Grand Unified Theory of Mind and Brain Part I: Space-Time Approach to Dynamic Connectomes of C. elegans and Human Brains*.

Arisaka, K. (2022b). *Grand Unified Theory of Mind and Brain Part II: Neural Holographic Tomography (NHT) and Holographic Ring Attractor Lattice (HAL)*.

Arisaka, K., & Blaisdell, A. (2022). *Grand Unified Theory of Mind and Brain Part III: Holographic Visual Perception of 3D Space and Shape*.

Bustanoby, I., Krupien, A., Bacani, M., Chandrashekar, P., & Gangal, A. (2022). *Visual Perception of 3D Space and Shape by 7D Frame Translation in Time Part II: 3D Space Perception with Holographic Depth*.

Caharel, S., Collet, K., & Rossion, B. (2015). The early visual encoding of a face (N170) is viewpoint-dependent: A parametric ERP-adaptation study. *Biological Psychology*, *106*, 18–27. https://doi.org/10.1016/j.biopsycho.2015.01.010

Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. A. J., & Friston, K. (2017). Distinct Top-down and Bottom-up Brain Connectivity During Visual Perception and Imagery. *Scientific Reports*, *7*(1), 5677. https://doi.org/10.1038/s41598-017-05888-8

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex (New York, N.Y.: 1991)*, *7*(2), 181–192. https://doi.org/10.1093/cercor/7.2.181

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature Neuroscience*, *8*(10), 1380–1385. https://doi.org/10.1038/nn1537

Favelle, S. K., Palmisano, S., & Avery, G. (2011). Face Viewpoint Effects about Three Axes: The Role of Configural and Featural Processing. *Perception*, *40*(7), 761–784. https://doi.org/10.1068/p6878

Favelle, S. K., Palmisano, S., & Maloney, R. T. (2007). Things are Looking up: Differential Decline in Face Recognition following Pitch and Yaw Rotation. *Perception*, *36*(9), 1334–1352. https://doi.org/10.1068/p5637

Favelle, S., & Palmisano, S. (2012). The Face Inversion Effect Following Pitch and Yaw Rotations: Investigating the Boundaries of Holistic Processing. *Frontiers in Psychology*, *3*. https://www.frontiersin.org/article/10.3389/fpsyg.2012.00563

Favelle, S., & Palmisano, S. (2018). View specific generalisation effects in face recognition: Front and yaw comparison views are better than pitch. *PLOS ONE*, *13*(12), e0209927. https://doi.org/10.1371/journal.pone.0209927

Felleman, D. J., & Van Essen, D. C. (1987). Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. *Journal of Neurophysiology*, *57*(4), 889–920. https://doi.org/10.1152/jn.1987.57.4.889

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Klimesch, W., Fellinger, R., & Freunberger, R. (2011). Alpha Oscillations and Early Stages of Visual Encoding. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00118

Kolster, H., Janssens, T., Orban, G. A., & Vanduffel, W. (2014). The Retinotopic Organization of Macaque Occipitotemporal Cortex Anterior to V4 and Caudoventral to the Middle Temporal (MT) Cluster. *Journal of Neuroscience*, *34*(31), 10168–10191. https://doi.org/10.1523/JNEUROSCI.3288-13.2014

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011

McLaughlin, D., Shapley, R., Shelley, M., & Wielaard, D. J. (2000). A neuronal network model of macaque primary visual cortex (V1): Orientation selectivity and dynamics in the input layer 4Cα. *Proceedings of the National Academy of Sciences*, *97*(14), 8087–8092. https://doi.org/10.1073/pnas.110135097

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Pessoa, L. (2014). Understanding brain networks and brain organization. *Physics of Life Reviews*, *11*(3), 400–435. <https://doi.org/10.1016/j.plrev.2014.03.005>

van Polanen V, Davare M. Interactions between dorsal and ventral streams for controlling skilled grasp. Neuropsychologia. 2015 Dec;79(Pt B):186-91. doi: 10.1016/j.neuropsychologia.2015.07.010. Epub 2015 Jul 11. PMID: 26169317; PMCID: PMC4678292.

Puce, A., Allison, T., & McCarthy, G. (1999). Electrophysiological Studies of Human Face Perception. III: Effects of Top-down Processing on Face-specific Potentials. *Cerebral Cortex*, *9*(5), 445–458. https://doi.org/10.1093/cercor/9.5.445

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research*, *51*(3), 333–342. https://doi.org/10.1016/j.visres.2010.11.014

Rossion, B. (2009). Distinguishing the cause and consequence of face inversion: The perceptual field hypothesis. *Acta Psychologica*, *132*(3), 300–312. https://doi.org/10.1016/j.actpsy.2009.08.002

*Stirling ESRC 3D Face database*. (2021, February 6). http://pics.stir.ac.uk/ESRC/

Ta, B., Silva, M. E. M. D. M. E., Smith, A., Bartlett, K., & Drost, A. (2022). *Visual Perception of 3D Space and Shape by 7D Frame Translation in Time Part III: 2D Shape Recognition by Log-scaling*.

Todd, J. T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences*, *8*(3), 115–121. https://doi.org/10.1016/j.tics.2004.01.006

Turati, C., Bulf, H., & Simion, F. (2008). Newborns’ face recognition over changes in viewpoint. *Cognition*, *106*(3), 1300–1321. https://doi.org/10.1016/j.cognition.2007.06.005

Zachariou, V., Klatzky, R., & Behrmann, M. (2014). Ventral and Dorsal Visual Stream Contributions to the Perception of Object Shape and Object Location. *Journal of Cognitive Neuroscience*, *26*(1), 189–209. https://doi.org/10.1162/jocn\_a\_00475

### Comparison Between Groups

Although each group took data independently with different experimental setups and protocols, all four groups achieved similar V-shape patterns for corresponding protocols in both intercept and slope (Figure 5). Hence, data from each group could be pooled and analyzed as a single combined group.

Chart, histogram

Description automatically generated

**Figure 5** Protocol plots showing group aggregate fits on a common axis.

### Internal Group A

Internal Group A took reaction time data through attractor vs. distractor tests. Gamma distribution modes for Roll and Yaw were near to or less than 1, providing strong support for a linear relationship between reaction time and rotational angle, while the mode for Pitch was higher but acceptable. Roll (n = 14) was nearly symmetric while Yaw (n = 14) was more symmetric, and Pitch (n = 12) was more asymmetric. As with the aggregate, Roll was the least steep and Pitch was the steepest protocol. After normalizing individual datasets about the aggregate via Cousineau normalization, Roll and Yaw appeared to have strong linear fits with small reduced-chi-squares and reasonable error bars. However, Pitch had a larger reduced chi-square, suggesting that the relationship may not be completely linear, especially at central angles where the data appears to curve like a parabola. The Pitch data was downshifted from the other protocols with a faster unrotated reaction time.

### Internal Group B

Internal Group B took reaction time data through 3-choice reaction time tests. As with group A, face data gamma distribution modes supported linear predictions for Roll and Yaw, while the mode for Pitch was higher. Both roll (n = 15) and yaw (n = 15) were slightly asymmetric with steeper negative slopes than positive slopes. Pitch (n = 15) was asymmetric in the opposite direction, with a steeper positive slope than the negative slope. Nonetheless, groups A and B agreed with Roll being the least steep and Pitch being the steepest. Roll had the strongest linear prediction with a reduced chi-square near one while Yaw and Pitch had large error bars at larger rotation angles due to a subset of participants having nonlinear relations between the reaction time and rotation or variation in linear slope between participants, leading to small reduced-chi-squares.

### External Group C

External Group C took reaction time data for only faces through shortened 3-choice reaction time tests. Gamma distribution modes supported linear predictions for all three protocols. Roll (n = 13) was symmetric and had the least steep slope. Like group B, yaw (n = 13) was slightly asymmetric with a steeper negative slope than the positive slope, and Pitch (n = 13) had the steepest slopes, which were asymmetric with a steeper positive slope than the negative slope. Except for data at -120°, roll had reasonable error bars and small reduced chi-squares, indicating a strong linear fit. On the other hand, yaw and pitch had very large error bars due to variation in slopes between participants, leading to decreased reduced chi-squares. Pitch appeared linear, but yaw appeared to curve upward from the linear prediction at larger rotations.

### External Group D

External Group D took reaction time data through a gamified version of the shortened 3-choice reaction time tests. Gamma distribution modes supported linear predictions for all three protocols, although Roll was the weakest. Roll (n = 14) was the least steep and was slightly asymmetric with a steeper positive slope than the negative slope while Yaw (n = 14) was symmetric with an intermediate slope. Pitch (n = 14) was the steepest and most asymmetric, also with a steeper positive slope than a negative slope. Unlike the other groups, Roll had the weakest linear prediction with large reduced chi-squares, indicating a nonlinear relation. Especially for positive angles, the data appeared to curve downwards at larger angles, with a smaller reaction time for full rotation than for intermediate rotations up to 120°. Yaw and Pitch appeared mostly linear but also had large reduced chi-squares. However, except for positive yaw, these values were reasonable given the small error bars. Positive yaw appeared linear excluding an increased reaction time at 90°.

## 2.2 Language Recognition