Perceiving the mean:

A study investigating whether the low-level visual cue spatial frequency adds to the computation of higher-level ensemble coding.

Teun Beumer

Master Thesis

16-7-2021

Abstract

Ensemble coding enables humans to quickly assess mean properties of visual cues. Researchers discriminate between low-level ensemble coding, enabling humans to quickly assess motion, direction, speed, colour, hue, facial expression, orientation, family resemblance or size of objects in an ensemble, and higher-level ensemble coding which enables humans to quickly assess the mean emotional state or gaze of groups of people. The mystery whether higher-level ensemble coding has a computational basis in low-level visual cues remains. There seems to be ample evidence suggesting spatial frequency playing a role in decoding emotional states. In this thesis it was investigated whether the low-level cue spatial frequency added to the computation of higher-level ensembles. Analyses showed no evidence to this effect suggesting high-level ensemble coding is primarily a feature based on holistic face perception.

Applied Cognitive Psychology
Utrecht University
Supervisor: dr. Sjoerd Stuit
Second supervisor: dr. Niilo Valtakari
Word count: 2645
ECTS: 27,5
Student number: 5954398
For further inquiries into this thesis don't hesitate to contact me on teun44b@gmail.com

Introduction

While walking through a forest, field or on a beach. One thing which becomes clear is that the world is full of the same things: leaves, branches, blades of grass, grains of sands, droplets of water. This redundancy gives rise to the question how the brain copes with all these impressions and details without overheating. Research into this very question dates back to early Gestalt Psychology (Kofka, 1935). However, this research really flourished in the next century with numerous researchers investigating the phenomenon of Ensemble Coding.

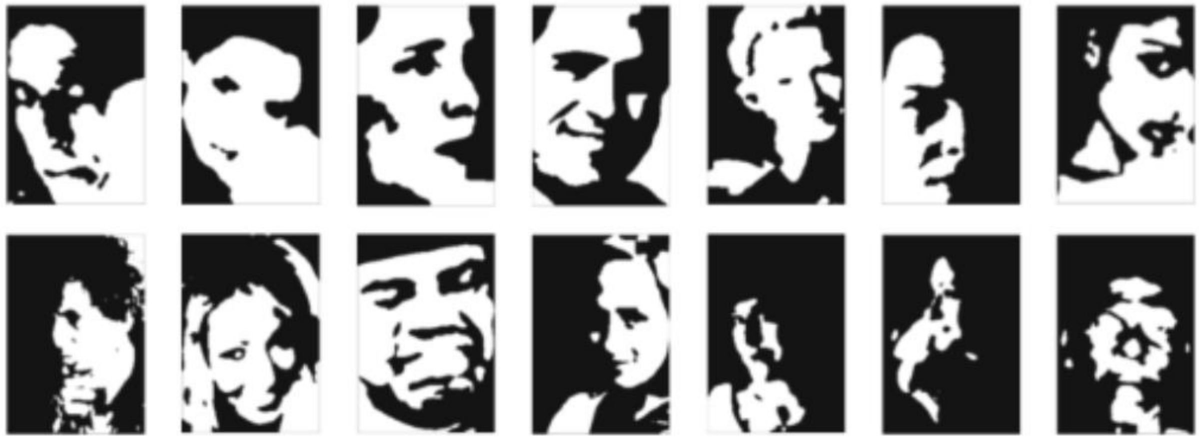
Where there is structure, there is redundancy, and were there is redundancy there is an opportunity to form a compressed and efficient representation of information (Alvarez, 2011, Haberman & Whitney, 2012). To take advantage of this structure and redundancy the brain represents these objects as ensembles instead of individuals: the brain sees the forest first then the trees (Cavanagh, 2001). This mechanism called ensemble coding derives statistical summary information from similar objects in the environment (Whitney, Yamanashi Leib, 2018). It makes it possible for humans to readily (within 100 milliseconds) report the motion, direction, speed, colour, hue, facial expression, orientation, family resemblance, gaze direction or size of objects in an ensemble (Whitney & Yamanashi Leib, 2018). Ensemble coding is so efficient, that humans can more easily report the average value of a parameter (e.g. colour) than judge whether an object was present (Potter, 1976; Rensink, Regan & Clark, 1997; Ariely, 2001).

Researchers discriminate between different levels of ensemble perception. Low-level ensemble perception has a computational basis in low-level cues, cues such as orientation, speed, hue, or size (Watamaniuk & McKee 1998,Watamaniuk et al. 1989). Ensemble coding seems to play a role in higher level features as well. Features such as emotion recognition, gaze direction and identity identification (Haberman & Whitney 2007, 2009). However there is some evidence which suggest two different ensemble coding mechanisms at work, one for low-level visual cues and one for higher-level cues (Haberman, Brady & Alvarez, 2015). How high-level ensemble statistics are computed remains a mystery (Li, Ji, Tong, Ren, Chen, Liu & Fu, 2016; Alvarez, 2011).

One theory suggests that high-level ensemble statistics are computed using both high- and low-level features, with these features both contributing to a holistic image: an image where the whole adds up to something that is more than its parts (Haberman & Whitney, 2010; Whitney, Haberman & Sweeny, 2014; Han, Leib, Chen & Whitney, 2020). For instance, ensemble perception of faces was negatively impacted by scrambling, a technique whereby holistic processing is impaired as it shows the parts instead of the whole (Haberman & Whitney, 2007; Sweeny, Haroz, & Whitney, 2013; Yamanashi Leib, Kosovicheva, & Whitney, 2016). However, Richler, Mack, Palmeri & Gauthier (2011) argued that alteration of photographs maybe only delays ensemble coding. And, furthermore, ensemble coding can still occur for inverted crowds (Elias, Dyer, & Sweeny, 2017; Sweeny & Whitney, 2014). To address the question whether ensemble coding also occurs for higher-level visual cues such as faces without part-based cues such as identifiable features or surface texture characteristics, Han, Yamanashi Leib, Chen & Whitney (2020) conducted experiments with Mooney faces. Mooney faces are black and white shadow defined images that cannot be recognized in a part-based manner, see Figure 1 (Han, Yamanashi Leib, Chen & Whitney, 2020). These researchers confirmed that integration of multiple Mooney faces into ensemble representations, concluding furthermore that ensemble perception functions when holistic information is maximized and that higher-level ensemble coding is more than averaging individual features.

Figure 1.

An example of Mooney faces used in the experiment of Han et al. (2020) described above.

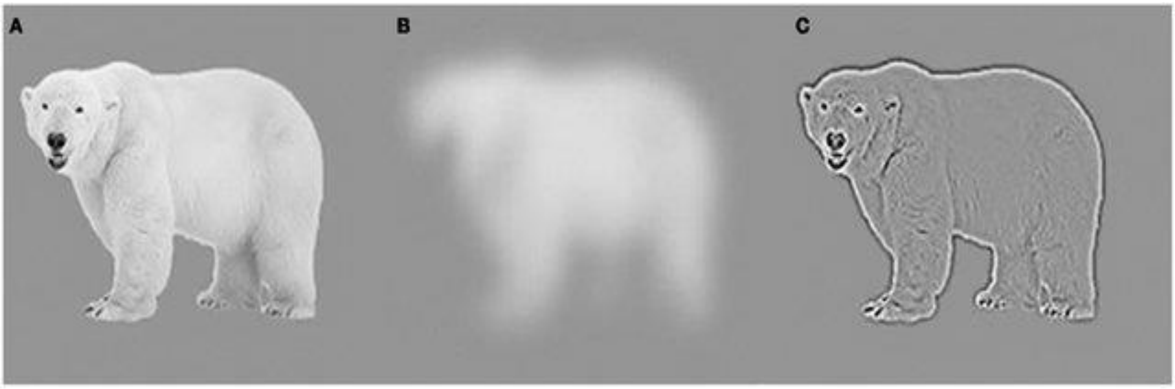


Note. From

But, it seems unlikely that high-level ensemble coding only works for holistic images. Numerous studies with stimuli containing both high- and low-level visual cues reported participants demonstrating ensemble coding (Leib, Puri, Fischer, Bentin, Whitney & Robertson, 2012; Rhodes, Neumann, Ewing, Bank, Read, Engfors, Emiechel & Palermo, 2018; Roberts, Cant & Nestor, 2019; Robson, Palermo, Jeffery & Neumann, 2018, Li, Ji, Tong, Ren, Chen, Liu & Fu, 2016). And having such a automatic and fast mental mechanism for computing ensemble statistics, which works even in people with impaired face perception (Unilateral Spatial Neglect), it makes hardly any sense that it solemnly relies on holistic representations.

Figure 2.

Showing a polar bear in both high and low spatial frequency (A); only in low frequency (B) and only in high spatial frequency (C).



Note. From *Predictive Feedback and Conscious Visual Experience* [image], by Panichello, Cheung & Moshe, 2013, *Frontiers in Psychology,* 3, p. 620.

To test whether high-level ensemble coding has any computational basis in low-level visual cues, this paper proposes to investigate Spatial Frequency as a low-level cue whereof higher-level ensemble statistics could be computed. Low-level ensemble statistics seem to be computed using spatial frequency (Oliva & Torralba, 2001; Torralba & Oliva, 2003; Alvarez, 2011; Kanaya, Hayashi, Whitney, 2018). Spatial frequency is a characteristic of luminance variations across space (Kumar & Srinivasan, 2011). High spatial frequency content provides local details from within a stimulus and low spatial frequency provides information about the global aspect of a stimulus (Kumar & Srinivasan, 2011). See Figure 2 for examples of high and low spatial frequency. Kumar & Srinivasan (2011) continue to conclude that higher spatial frequencies are linked to the identification of a sad expression and low spatial frequencies are linked to identification of a happy expression. Furthermore, Stuit, Kootstra, Terburg, van den Boomen, van der Smagt, Kenemans & van der Stigchel (2021) found spatial frequency to be a good predictor of initial eye movement between two faces. Additionally, Awasthi, Friedman & Williams (2011) found that patients with prosopagnosia (the inability to recognize faces) showed a 230ms delay in low spatial frequency processing as compared to controls, suggesting that processing low spatial frequency information is critical for the development of normal face perception. On top of this, the field of machine learning has been incorporating spatial frequency information for training deep-learning algorithms to classify images based on emotions, age or gender with relative success as opposed to other methods of classification (Lee, Gu Kim, Kim & Ro, 2018).

Prior research has shown that humans are capable of computing average statistics, better known as ensemble statistics. Humans compute low-level ensemble statistics for e.g. orientation, color, shape and high-level ensemble statistics for e.g. faces. The exact visual stimuli and mechanisms exploited to compute these ensembles remains however, unknown. There seems to be ample evidence pointing towards the low-level visual cue spatial frequency as being involved in the computation of higher level ensemble statistics. This research will investigate whether high-level ensemble statistics have any computational basis in low-level visual cues. Does high-level ensemble coding have a computational basis in low-level visual cues such as spatial frequency?

Methods

Participants

49 people participated in the current study. Prior to the experiment participants filled in a consent form informing the them their data would be anonymised. The study was approved by the ethical committee of the Utrecht University.

Apparatus

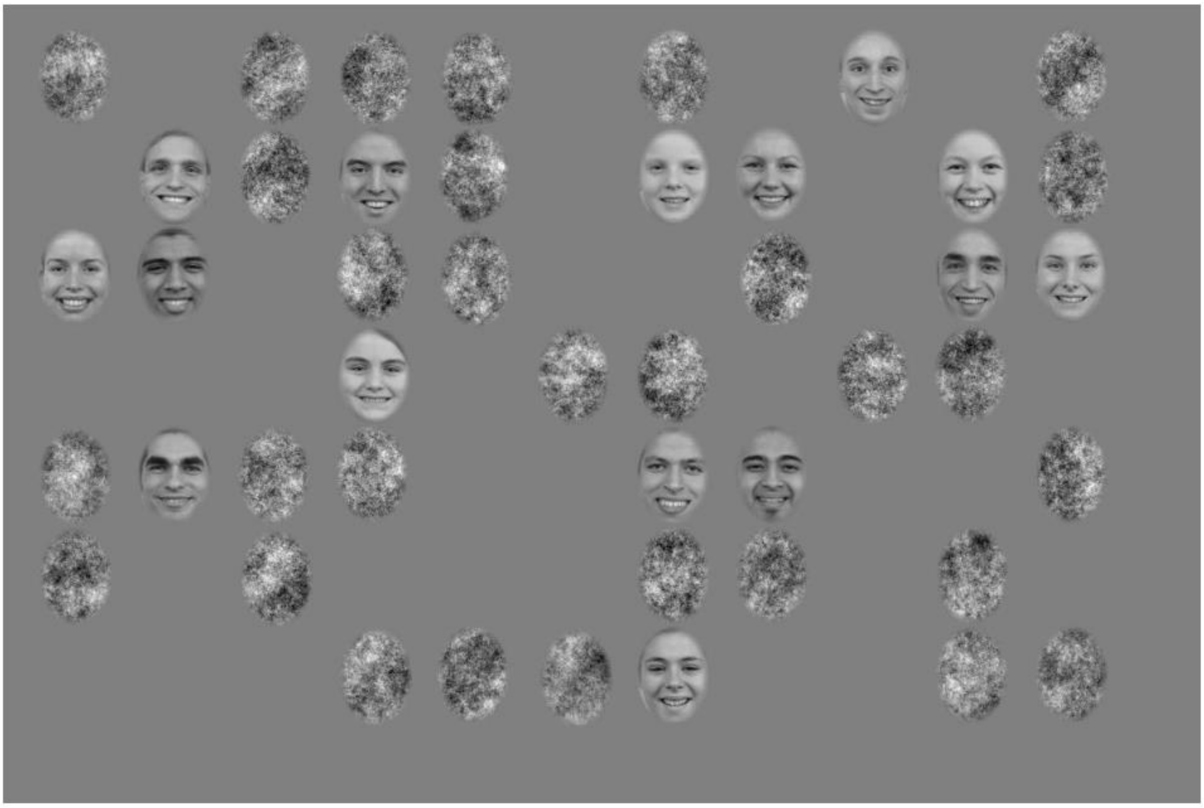
Stimuli were created with an Apple Mac Pro computer running OS X and Matlab 2019b with the psychophysics Toolbox extensions version 3.0. The experiment was coded in Inquisit 6. The experiment ran in InquisitPlayer, each participant was instructed on how to install it on their PC, the experiment could only be performed on a PC. After the experiment an explanation of how to uninstall the InquisitPlayer was shown.

Stimuli

The stimuli consisted of 480 displays. These displays contained 15 faces which could range from 0 angry or happy faces to 15 angry or happy faces, see Figure 2 for an example. Furthermore, next to face images the displays contained 30 noise patches. Depending on the condition, the noise patches were created by combining the Fourier magnitude spectrum of either happy faces or angry faces with a random phase spectrum. Each noise patch was based on a unique face. A third condition of noise patches consisted of 1/f noise patches. These faces were sampled from the Radboud Faces database. Only faces with a frontal gaze were used and all faces were converted to greyscale.

Figure 3.

An example of a display used in the experiment, the display contains 15 happy faces along with 30 noise patches containing the spatial frequency of either happy, angry or 1/f noise. Each display was shown for 100ms.



Design & Procedure

Participants received written instructions on how to perform the experiment. Prior to the trials the participants were asked their age and sex. After that the participants did 10 practice trials whereafter the experiment began. Participants were shown a white cross for 1000ms in the middle of their display as a prompt, after that the display was shown for 100ms followed by a black display with an instruction reminder asking them whether they perceived the display as more angry or happy. Participants performed a two alternate choice task after the display was shown, they were instructed to press the “a” button if they thought the displays was more angry and they were instructed to press the “h” button if they thought the displays was more happy. The instruction reminder stayed on the screen until a response was recorded. Participants could stop at anytime by pressing “*ctrl* + q”.

Analyses

To establish whether there was any difference between the white-noise, happy-noise and angry-noise conditions Points of Subjective Equality (PSE’s) were calculated. The PSE is the required number of angry faces in a display needed for a participant to give an angry response 50% of the time as estimated by a fitted regression model. This linear regression model was fitted over the data of each participant and each condition. Analyses were conducted in JASP and SPSS.

Results

To test whether the participants demonstrated an ensemble effect. Using a one-sample t test the slope of the regressions were established to be significantly bigger than 0. The means and standard deviations of the slopes for the white noise, happy noise and angry noise are respectively: M = 0.051, SD = 0.012; M = 0.051, SD = 0.013; M = 0.052, SD = 0.013. The t tests of the white noise, happy noise and angry noise are respectively: t(48) = 29.760, p < .001, d = 4.251; t(48) = 27.985, p < .001, d = 3.998; t(48) = 28.916, p < .001, d = 4.131. See figure 3.

Figure 4.

A bar graph showing the means and standard deviations of the Points of Subjective Equality (PSE's) per condition.

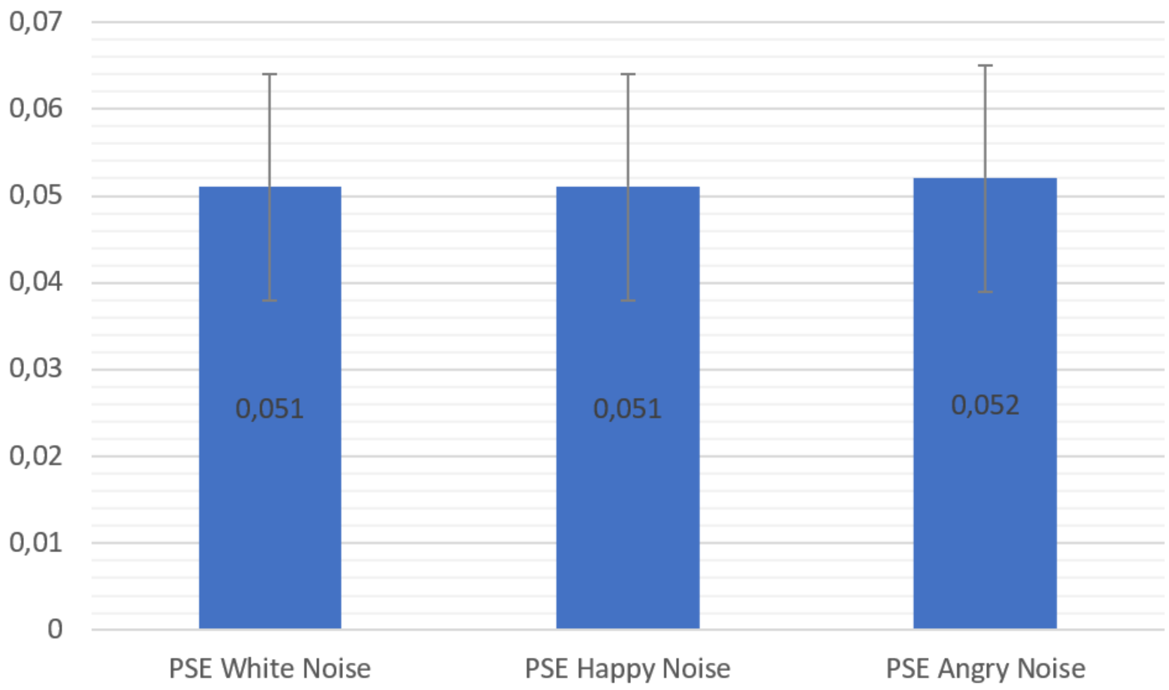
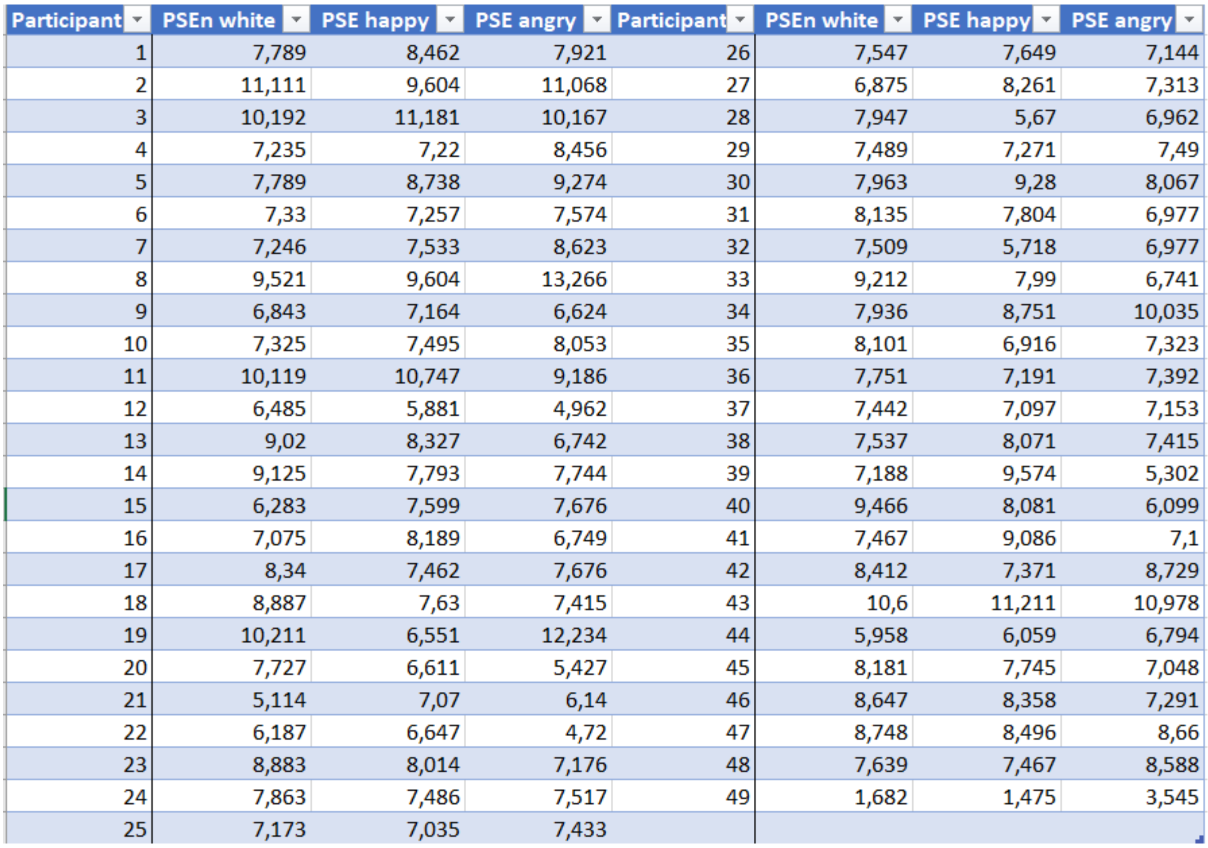


Figure 5.

A table showing the mean PSE's for each participant for each spatial frequency noise type.

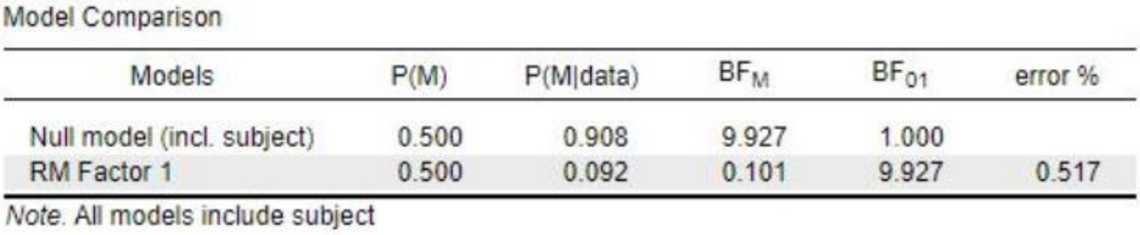


Participant	PSEn white	PSE happy	PSE angry	Participant	PSEn white	PSE happy	PSE angry
1	7,789	8,462	7,921	26	7,547	7,649	7,144
2	11,111	9,604	11,068	27	6,875	8,261	7,313
3	10,192	11,181	10,167	28	7,947	5,67	6,962
4	7,235	7,22	8,456	29	7,489	7,271	7,49
5	7,789	8,738	9,274	30	7,963	9,28	8,067
6	7,33	7,257	7,574	31	8,135	7,804	6,977
7	7,246	7,533	8,623	32	7,509	5,718	6,977
8	9,521	9,604	13,266	33	9,212	7,99	6,741
9	6,843	7,164	6,624	34	7,936	8,751	10,035
10	7,325	7,495	8,053	35	8,101	6,916	7,323
11	10,119	10,747	9,186	36	7,751	7,191	7,392
12	6,485	5,881	4,962	37	7,442	7,097	7,153
13	9,02	8,327	6,742	38	7,537	8,071	7,415
14	9,125	7,793	7,744	39	7,188	9,574	5,302
15	6,283	7,599	7,676	40	9,466	8,081	6,099
16	7,075	8,189	6,749	41	7,467	9,086	7,1
17	8,34	7,462	7,676	42	8,412	7,371	8,729
18	8,887	7,63	7,415	43	10,6	11,211	10,978
19	10,211	6,551	12,234	44	5,958	6,059	6,794
20	7,727	6,611	5,427	45	8,181	7,745	7,048
21	5,114	7,07	6,14	46	8,647	8,358	7,291
22	6,187	6,647	4,72	47	8,748	8,496	8,66
23	8,883	8,014	7,176	48	7,639	7,467	8,588
24	7,863	7,486	7,517	49	1,682	1,475	3,545
25	7,173	7,035	7,433				

To test whether the spatial frequency noise types had an effect on ensemble coding of the faces in the display. The differences between the points of subjective equality (PSE's) of the different noise types were investigated using a Bayesian repeated measures analysis of variance (ANOVA). The Bayesian ANOVA returned a BF of 0,908 in favour of the null-hypotheses. Stating there is more evidence for the null-hypothesis than for the model. Mauchly's test indicated the assumption of Sphericity was not violated. See Figure 6.

Figure 6.

Bayesian ANOVA output table.



Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{01}	error %
Null model (incl. subject)	0.500	0.908	9.927	1.000	
RM Factor 1	0.500	0.092	0.101	9.927	0.517

Note. All models include subject

Discussion

In this thesis the question whether high-level ensemble coding has a computational basis in low-level visual cues was investigated. People almost instantly recognize the average expressions on groups of faces, general scenes and mean directions of objects via ensemble coding (Haberman & Whitney, 2018). This paper found evidence adding to the validity of ensemble coding. Ensemble coding enables humans to perceive the mean, however, which information plays a role in assessing the mean emotional expression of a group of faces is quite unclear (Li, Ji, Tong, Ren, Chen, Liu & Fu, 2016; Alvarez, 2011). In this thesis it was hypothesized that spatial frequency is part of the computational basis of high-level ensemble coding. Based on the evidence provided in the Bayesian ANOVA, spatial frequency does not seem to influence the perception of the faces in the displays.

The mystery of how high-level ensemble statistics are computed still remains. In this experiment 15 faces ranging from happy to sad were shown for 100ms. The more angry faces shown the more angry responses were recorded. Prior research showed that when there is little time (around 50ms) there can be no precise individual representations, the brain then computes average statistics instead (Pavlovskaya, Bonneh, Soroker, & Hochstein, 2010; Yamanashi Leib, Landau, Baek, & Chong, 2012; Yamanashi Leib, Puri, Fischer, Bentin, Whitney & Robertson, 2012). It might be possible that when the stimuli were only shown for 100ms the brain prioritizes to scan the 15 faces first and afterwards the spatial frequency noise patches. Maybe the brain ran out of time and did not register the spatial frequency noise patches correctly. This could explain why there is an ensemble effect but there is no influence on the ensemble effect from the noise patches.

Another explanation for the data might have to do with the emotions chosen in the trials. Kumar and Srinivasan (2011) found that happy expressions were identified faster than sad expressions. This was also found by other researchers (Alves et al., 2009; Eastwood et al., 2001; Hitenan & Leppanen, 2004; Kirita & Endo, 1995; Srivastava & Srinivasan, 2010). In this research it was hypothesized that the display containing an incongruent situation, e.g. angry faces with happy spatial frequency noise would take more angry faces on average to record an angry response than a display containing an congruent situation, e.g. angry faces with angry spatial frequency noise. However the tendency to identify happy faces quicker does not show in this data. If high-level ensemble coding has a computational basis in low-level visual cues such as spatial frequency, the incongruent displays would have taken more angry faces for a participant to categorize it as angry. This adds to the theory that high-level ensemble coding is primarily based on holistic stimuli and has little basis in lower-level visual cues (Han, Yamanashi Leib, Chen & Whitney, 2020).

Evidence for primarily holistic perception of faces was provided by Han, Yamanashi Leib, Chen and Whitney (2020) when they showed that ensemble coding also occurs if participants were shown Mooney faces. Mooney faces, consisting only of black and white shadow based tones, can not be recognized in a part based manner (Han, Yamanashi Leib, Chen & Whitney, 2020). These findings show ensemble perception to be a broadly useful tactic to quickly summarize any level of visual representation (Han et al., 2020). They give however, little insight into whether there are different mechanisms at work for both high- and low-level ensemble coding.

It might very well be that there is one primary mechanisms at work which takes in information from both high- and low-level visual cues, on the other hand the theory of two different mechanisms: one for high-level ensemble coding and one for low-level ensemble coding is also possible (Haberman, Brady & Alvarez, 2015; Neumann, Ng, Rhodes, &Palermo, 2017; Li, Ji, Tong, Ren, Chen, Liu &Fu, 2016). Faces are holistic stimuli, and centuries of research into the human brain have already established the fusiform gyrus as having the solemn purpose of recognizing faces (Lopatina, Komleva, Gorina, Higashida & Salmina, 2018). As recognizing and conveying emotional states provides adaptational value, recognizing the emotions of an entire group might provide that value as well, resulting in there being a structure solemnly responsible for identifying higher-level ensembles (Parr, Winslow, Hopkins & de Waal, 2000). This remains, for the time being, only speculation. A lot more research is needed into the computational and biological basis of higher-level ensemble coding. It might be interesting to do EEG research on the fusiform gyrus of participants assessing emotional states of Mooney faces to give further insights into the bases of high-level ensemble coding.

This thesis investigated whether there is a computational basis for higher-level ensemble statistics in the low-level visual cue spatial frequency. Evidence for ensemble statistics was found, however spatial frequency seems not to influence these high-level ensemble statistics.

Literature

Alvarez, G.A. (2011). Representing multiple object as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122-131.

Awasthi, B., Friedman, J., & Williams, M. A. (2011). Faster, stronger, lateralized: low spatial frequency information supports face processing. *Neuropsychologia*, *49*(13), 3583-3590.

Cavanagh, P. (2001). Seeing the forest but not the trees. *Nature Neuroscience*, 4(7), 673–674.

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. *Perception & psychophysics*, *63*(6), 1004-1013.

Elias, E., Dyer, M., & Sweeny, T. D. (2017). Ensemble perception of dynamic emotional groups. *Psychological Science*, *28*(2), 193-203.

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of vision*, *9*(11), 1-1.

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. *From perception to consciousness: Searching with Anne Treisman*, 339-349.

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718.

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics*, *72*(7), 1825 1838.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current biology*, *17*(17), R751-R753.

Han, L., Leib, A. Y., Chen, Z., & Whitney, D. (2020). Holistic ensemble perception. *Attention, Perception, & Psychophysics*, 1-16.

Kanaya, S., Hayashi, M. J., & Whitney, D. (2018). Exaggerated groups: Amplification in ensemble coding of temporal and spatial features. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1879), 20172770.

Kirita, T., & Endo, M. (1995). Happy face advantage in recognizing facial expressions. *Acta psychologica*, *89*(2), 149-163.

Koffka, K. (1922). Perception: an introduction to the Gestalt-Theorie. Psychological bulletin, 19(10), 531.

Kumar, D. & Srinivasan, N. Emotion perception is mediated by spatial frequency content. *Emotion* **11**(5), 1144 (2011).

Lee, S. S., Kim, H. G., Kim, K., & Ro, Y. M. (2018, October). Adversarial spatial frequency domain critic learning for age and gender classification. In *2018 25th IEEE International Conference on Image Processing (ICIP)* (pp. 2032-2036). IEEE.

Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C. H., & Fu, X. (2016). Processing of individual items during ensemble coding of facial expressions. *Frontiers in psychology*, *7*, 1332.

Oliva, A., Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3), 145-175

Pavlovskaya, M., Bonneh, Y., Soroker, N., & Hochstein, S. (2010). Processing visual scene statistical properties in patients with unilateral spatial neglect. *Journal of Vision*, *10*(7), 280-280.

Rhodes, G., Neumann, M., Ewing, L., Bank, S., Read, A., Engfors, L. M., ... & Palermo, R. (2018). Ensemble coding of faces occurs in children and develops dissociably from coding of individual faces. *Developmental Science*, *21*(2), e12540.

Rhodes G, Neumann MF, Ewing L, Palermo R. 2019. Reduced set averaging of face identity in children and adolescents with autism. Q. J. Exp. Psychol. 68(7):1391–403

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision research*, *51*(3), 333-342.

Robson, M.K., Palermo, R., Jeffery, L., Neumann, M.F. (2018). Ensemble coding of face identity is present but weaker in congenital prosopagnosia. Neuropsychologia, 111, 377-386.

Srivastava, P., & Srinivasan, N. (2010). Time course of visual attention with emotional faces. *Attention, Perception, & Psychophysics*, *72*(2), 369-377.

Stuit, S. M., Kootstra, T. M., Terburg, D., van den Boomen, C., van der Smagt, M. J., Kenemans, J. L., & Van der Stigchel, S. (2021). The image features of emotional faces that predict the initial eye movement to a face. *Scientific Reports*, *11*(1), 1-14.

Sweeny TD, Haroz S, Whitney D. 2013. Perceiving group behavior: sensitive ensemble coding mechanisms for biological motion of human crowds. J. Exp. Psychol. Hum. Percept. Perform. 39(2):329–37

Sweeny TD, Whitney D. 2014. Perceiving crowd attention: ensemble perception of a crowd’s gaze. Psychol. Sci. 25(10):1903–13

Watamaniuk SN, McKee SP. 1998. Simultaneous encoding of direction at a local and global scale. Percept. Psychophys. 60(2):191–200

Watamaniuk SN, Sekuler R, Williams DW. 1989. Direction perception in complex dynamic displays: the integration of direction information. Vis. Res. 29(1):47–59

Whitney, D., Haberman, J., & Sweeny, T. D. (2014). 49 From Textures to Crowds: Multiple Levels of Summary Statistical Perception.

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual review of psychology*, *69*, 105-129.

Yamanashi Leib A, Kosovicheva A, Whitney D. 2016. Fast ensemble representations for abstract visual impressions. Nat. Commun. 7:13186

Yamanashi Leib A, Landau AN, Baek Y, Chong SC, Robertson L. 2012a. Extracting the mean size across the visual field in patients with mild, chronic unilateral neglect. Front. Hum. Neurosci. 6:267

Yamanashi Leib A, Puri AM, Fischer J, Bentin S, Whitney D, Robertson L. 2012b. Crowd perception in prosopagnosia. Neuropsychologia 50(7):1698–707