Seeing faces in the noise: Stochastic activity in perceptual regions of the brain may influence the perception of ambiguous stimuli

Heather A. Wild

Thomas A. Busey

Indiana University, Bloomington

Address Correspondence to:

Heather Wild (hwild@indiana.edu)

Thomas Busey (busey@indiana.edu)

Abstract

Research on binocular rivalry and motion direction discrimination suggests that stochastic activity early in visual processing influences the perception of ambiguous stimuli. Here we extend this to higher-level tasks of word and face processing. Experiment 1 used blocked gender and word discrimination tasks and Experiment 2 used a face versus word discrimination task. Stimuli were embedded in noise and some trials contained only noise. In Experiment 1, we found a larger response in the N170, an ERP component associated with faces, to the noise-alone stimulus when observers were performing the gender discrimination task. The noise-alone trials in Experiment 2 were binned according to the observer's behavioral response and there was a greater N170 when they reported seeing a face. After considering various top-down and priming related explanations, we raise the possibility that seeing a face in noise may result from greater stochastic activity in neural face processing regions.

A basic goal of cognitive neuroscience is to link behavior with neural mechanisms. Two notable successes come from research on binocular rivalry and motion direction discrimination, in which an ambiguous stimulus is presented and physiological correlates are found between the reported percept and ongoing activity in visual areas such as V1 and MT (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996; Britten, Shadlen, Newsome, & Movshon, 1993) as well as extrastriate areas (Tong, Nakayama, Vaughan & Kanwisher, 1998). In the present work we seek to generalize this principle to the domains of word and face processing, which also may involve specialized neural areas in the inferotemporal cortex (Kanwisher, Stanley, & Harris, 1999, but also see Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999 for evidence of expertise effects in the same area). While prior work has been done with single-cell and fMRI recording, here we rely on known response properties of electrophysiological (EEG) measures.

In the present study, we address the relation between face-related brain activity and the reported percept by embedding faces and words in noise and using noise-alone displays to create an ambiguous stimulus. We will rely on the N170 component of the event-related potential[1], which has been linked to activity in face-related regions of the human. Jeffreys (1989) showed that Mooney faces elicit a negative-going component that occurs 170 ms after stimulus onset. However, when these faces are inverted, they are difficult to interpret as a face and the N170 is likewise attenuated. Numerous subsequent studies also found that faces elicit a strong N170 component (Bentin, 1998; Bentin, Allison, Puce, Perez, & McCarthy, 1996; Olivares & Iglesias, 2000). This downward deflection is largest over the temporal lobes (Bentin et al., 1996). While there is some

---

[1] The event related potential (ERP) averages the EEG response over trials.

disagreement as to the precise neural locus of the N170, its latency and spatial location

suggest that it represents activity in regions typically associated with early perceptual

processing of faces and other complex visual stimuli.

In addition to these bottom-up factors, perceptual expertise and context modulate

the N170. Tanaka and Curran (2001) found a larger N170 in bird and dog experts for

faces of animals for which an individual was an expert. Rossion et al. (2002) trained

observers to individuate novel objects called 'greebles,' and found expertise effects in the

N170. To demonstrate contextual effects, Bentin and colleagues (Bentin & Golland,

2002; Bentin, Sagiv, Mecklinger, Frederici, & von Cramon, 2002) presented pairs of dots

to observers which evoked only a weak N170 response. The dots were subsequently

shown surrounded by face features that made them interpretable as eyes. After priming

with a face context, the N170 to the dots alone increased. These studies demonstrate that

context interacts with perceptual information to modulate the N170.

In the present study, we examine the link between the activity reflected by the

N170 and the behavioral response in a task that has been made ambiguous with respect to

faces and words. We recorded EEG from observers while they viewed faces and words

embedded in random pixel noise, and some trials contained only the noise as an

ambiguous stimulus. In the first experiment, observers completed a word discrimination

task ('honesty' versus 'trust') and a gender categorization task. This extends prior work

on contextual influences on the N170 by examining the effect of observer expectations on

the EEG response to a stimulus without readily interpretable face-like features (i.e., the

noise-alone display). In Experiment 2, we presented the same stimuli to observers, but

intermixed the trials so that observers made a face versus word judgment on each trial.

Thus, observers had no reason to expect a face versus a word on any particular trial. The question is whether we see modulation of the N170 without the influence of expectations and bottom-up perceptual information.

Experiment 1

Participants

Nine right-handed observers participated in the study. These observers were students at IU whose participation comprised part of their labwork or coursework. All observers were naïve as to the purpose of the study.

Apparatus

The EEG was sampled at 1000 Hz and amplified by a factor of 20,000 (Grass amps model P511K) and band-pass filtered at .1 - 100 Hz (notch at 60 Hz). Signals were recorded from sites F3, F4, Cz, T5, and T6, with a nose reference and forehead ground; all channels had below 5 kOhm impedance. Recording was done inside a Faraday cage. Eyeblink trials were identified from a characteristic signal in channels F3 and F4 and removed from the analysis with the help of blink calibration trials. Images were shown on a 21 inch (53.34 cm) Macintosh color monitor approximately 44 inches (112 cm) from participants.

Stimuli

The entire stimulus set appears in Figure 1. Face stimuli consisted of grayscale frontal views of one male and one female face with neutral expressions, generated using Poser[TM] (Metacreations). Faces subtended a visual angle of 2.1 x 2.8 degrees. Two low-imagery words were chosen for the second task ("Honesty" and "Trust"). Words subtended a visual angle of 1.1 x .37 degrees. All stimuli were embedded in white noise (4.33 x 4.33

degrees of visual angle) that was identical (i.e. not resampled) on all trials. This single

noise field was uniform white noise with a mean luminance of 30 cd/ m^2, and a standard

deviation of 14.5 cd/m^2. This noise was added to the faces and words on a pixel-by-pixel

basis. The faces and words had standard deviations of 7.0 and 1.7 cd/m^2 at full contrast

prior to the addition of the noise. To create the low-contrast versions, the contrast of the

faces and words was adjusted until independent raters judged that they were near

threshold and approximately equally-detectable. The standard deviations of the low

contrast faces and words were 1.27 and .52 cd/m^2 respectively.

Procedure

Observers completed blocks of trials for face discrimination and word

discrimination tasks. Observers freely viewed the stimulus, and although no fixation

point was used the stimulus appeared in the same location on each trial and was framed

by the edge of the monitor. For the gender discrimination task, observers had to indicate

whether the face was male or female, and for the word discrimination task, observers had

to indicate whether the word was 'honesty' or 'trust.' They were told that there was a

stimulus present on every trial, despite the fact that one-third of the trials were noise-

alone. Observers were also told that faces and words would appear equally often.

Stimuli were presented for 1000 ms. EEG was recorded from 100 ms prior to

stimulus onset to 1100 ms post-stimulus onset. There were 100 trials with a word or a

face at each contrast level and 200 noise-alone trials for a total of 600 trials. Observers

responded after each trial via a numeric keypad.

Results and Discussion

The data from Experiment 1 is shown in Figure 2. Consider first the thin lines, which are the responses to the high contrast faces and words. The amplitude of the N170 is greater for high contrast faces than for high contrast words for both electrode sites. The thick light-grey lines, corresponding to the low contrast condition, show this same pattern. These data show that N170 differentiates between faces and words. The dark curves, highlighted in the lower panel of Figure 2, come from the noise-alone trials. The solid and dashed thick curves comes from blocks in which the observer expected a face or a word, respectively. In both channels the amplitude[2] of the N170 is significantly greater for noise-alone trials when observers are expecting a face rather than a word for electrode sites T5 (paired two-tailed $\underline{t}(8) = 2.62$, $\underline{p} < .05$) and T6 ($\underline{t}(8) = 2.35$, $\underline{p} < .05$).

These results extend the findings of Bentin's studies (Bentin & Golland, 2002; Bentin et al., 2002) to stimuli that contain no face-like features. Most importantly, the results of Experiment 1 show that the N170 can be modulated by the task of looking for a face, and not just by the physical presence of a face or face-like features. The next step is to use the N170 to address how activity might be related to a behavioral response when we remove contextual information as well. This was the aim of Experiment 2, which is identical to Experiment 1 except that we used a mixed design and had observers complete a face versus word discrimination task. The central question of Experiment 2 is whether, on noise-alone trials, observers will produce a larger N170 when they report seeing a face.

Experiment 2

Participants

---

[2] This was computed by taking the average amplitude in the time window from 140-200

Ten right-handed observers participated in the study.

Procedure

All stimuli in Figure 1 were presented in random order, and observers had to indicate whether a face or a word was embedded in the noise. They were told that there was a stimulus present on every trial, despite the fact that one-third of the trials were noise-alone. Observers were also told that faces and words would appear equally often. Observers responded via a joystick using a single finger, and were asked to make speeded responses, which was intended to eliminate additional guessing strategies not tied to the initial perceptual processing of the stimulus. There were 120 trials with a word or a face at each contrast level and 240 noise-alone trials for a total of 720 trials.

Results and discussion

EEG signals were averaged across trials for each subject based on the stimulus, and the noise-only trials were binned according to the subject's response (either 'face' or 'word'). Figure 3 shows the data for Experiment 2. The thin curves plot the data for the high-contrast faces and words. As in Experiment 1, we found a larger N170 for the high contrast face than the high contrast word.

The data that bear on the central question of the experiment come from the trials where only noise was presented, because on these trials the physical stimulus is held constant and only the response of the observer changes. These data are plotted as thick lines in the lower panel of Figure 3. As shown in the lower right panel of Figure 3, the N170 at the right temporal channel (T6) associated with a 'face' response to the noise-

---

ms for each condition.

alone stimulus is significantly larger than the N170 associated with a 'word' response (two-tailed, $\underline{t}(9) = 2.74$, $\underline{p} < .05$).

For the left temporal channel (T5), shown in the lower left panel of Figure 3, this difference is present but not significant, $\underline{t}(9) = 1.54$, ns. Note that in T5 the difference between the N170 amplitudes for the high contrast words and faces is much smaller than in channel T6. This is consistent with other right-hemisphere laterality effects involving faces (Farah, 1990).

We also analyzed the P100 and P300 components by averaging the amplitudes in the 80-130 ms and 260-340 ms windows and found no significant differences for either channel, nor in the other 3 channels. Thus the differences in the ERPs between 'face' and 'word' responses to the noise are confined to the right temporal lobe at about 170 ms after stimulus onset.

Observers were extremely accurate at discriminating high (mean = 99%) and low (mean = 97%) contrast faces and words. As seen in Table 1, there was a modest bias to say 'face' to the noise-alone trials such that observers made this response 62% of the time. Since there was a wide range in bias among the observers (11-97%), we examined whether the effect seen in the N170 was related to this bias. Effect size was defined as the difference of the average amplitude in the 150-200 ms window for word versus face responses. These data appear in Table 1; note that the effect is present for nine out of ten observers. Bias was modestly correlated with effect size ($r^2 = .40$, $\underline{p} < .05$). However, further analyses show that this correlation is driven by three observers with strong biases. When we remove these observers the correlation is no longer significant ($r^2 = .10$), but the difference in the N170 for word versus face response trials is still significant, $\underline{t}(6) =$

2.69, $\underline{p}$ < .05. Thus while there appears to be some individual differences in the response properties of the perceptual regions as indexed by the N170, the main results cannot be completely attributed to observer bias to say face.

The averages of observers' median reaction times for the different conditions appear in Table 2. Faces and words differ on many dimensions and this may have contributed to the RT differences at high and low contrast. However, the noise-alone trials contain the same stimulus which makes the comparison between the word-response and face-response trials reasonable. For these trials, there is no difference in reaction times.

The N170 occurs too early to simply be a signature of the observer's response after it had been executed; thus, while it is possible that the N170 neurons influence the decision, it is unlikely that the subject's decision influences the electrophysiological response at 170 ms. However, possible pre-trial influences also exist, such as priming from the previous trial, or perhaps expectations that a particular stimulus is going to appear (e.g., the Gambler's fallacy). We explored this possibility by examining whether the presentation of a face on the previous trial results in a larger N170 on current noise-alone trials. Figure 4 shows the ERPs binned according to each possible response to noise alone trials (i.e., 'face' or word'). These are also conditioned on whether the stimulus presented on the previous trial was a face or a word, such that there are four ERP traces shown. Consider first the two thick curves, which represent trials in which the observer responded 'face'. There is clearly no effect of the prior-trial stimulus, since the two curves are almost identical throughout the time period of interest (140-200 ms).

The thin curves in Figure 4 correspond to trials in which the observer responded 'word'. The difference in the amplitude between 140-200 ms is significant, $\underline{t}(9) = 3.24$, $\underline{p} = .01$. However, the differences occur late in the window, and are small compared with the overall main result, which can be recovered by comparing the average of the thin lines with the average of the thick lines. Furthermore, research shows that there appears to be little effect of prior exposure of the face on the N170 response (Cauquil, Edmonds, & Taylor, 2000). Given that the effects of the prior trial stimulus are small relative to the overall effects and are limited to trials in which the observer responds 'word,' we rule out the prior-trial priming hypothesis as a major explanation of the results.

General Discussion

In the present studies we established a relationship between face-related activity as indexed by the N170 and the eventual behavioral response. Experiment 1 allowed us to extend the work of Bentin and colleagues by eliminating structured face-like features from the stimulus and examining the influence of observer expectations on the N170. In Experiment 2, we found that observers show a greater N170 when they think they see a face in the noise rather than a word, an effect which is localized to the right temporal lobe around 170 ms post stimulus onset. This result cannot be linked to observer expectations as in Experiment 1 because we intermixed face and word trials, and we have eliminated any possible stimulus-driven influences because the conditions we are comparing use trials that contain identical noise. What might produce this result in the N170? Having ruled out explanations based on prior-trial priming and response bias, we consider other possible top-down influences such as fluctuations in attention to different types of information in the image, as well as mechanisms not related to top-down or bottom-up

processes, such as stochastic fluctuations in neural activity in the perceptual regions. Both

interpretations are interesting, and below we evaluate the evidence for each possibility.

The observer's decision may be influenced by the nature of the information that is

acquired, perhaps through tuning of spatial frequency channels or attention to different

face-like features in the noise. Faces tend to have lower spatial frequencies than words,

and observers may tune their spatial frequency filters to one range or another on a given

trial. Neurons that respond to faces may receive input from more cells that are tuned to

lower spatial frequencies and have a larger receptive field, and this may provide a

stronger response at the N170 if observers attend to lower spatial frequencies on noise-

alone trials. Evidence that the N170 is sensitive to spatial frequency information comes

from Goffaux, Gauthier, and Rossion (2003) who recorded EEG to low- and high-pass

filtered faces and cars. They found that the stronger N170 response for faces is specific to

low-pass filtered stimuli.

There are several pieces of evidence that argue against an explanation of our

effects based on dynamic tuning of spatial frequency filters. First, there is no evidence of

a preferential response to low-pass filtered faces during intracranial recording (McCarthy,

Puce, Belger, & Allison, 1999) and numerous studies show that line drawings of faces

elicit an N170 just as strongly as a real face (e.g., Bentin et al., 2002). Second, studies

have shown that attention to specific spatial frequency gratings causes increased

positivity in the ERP in the 100-200 ms range (i.e., around the N170), and this effect is

the same regardless of the spatial frequency of the attended stimulus (Bass, Kenemans, &

Mangun, 2002; Martinez, Di Russo, Anllo-Vento, & Hillyard, 2001). Thus, attending to

low spatial frequency bands does not automatically produce a larger N170. Third,

observers in the experiment had no motivation to alter their spatial frequency tuning on a trial-by-trial basis. Thus while we cannot rule out all possible strategies that observers may employ, we have controlled for as many as we can by using a mixed design and eliminated others by looking at prior-trial priming hypothesis.

Having ruled out as many top-down and prior-trial hypotheses as possible, we would like to raise the intriguing possibility that the behavioral response is directly related to the activity in the N170 neurons. The timing and the spatial location of the N170, as well as the tasks that modulate it, all suggest that the N170 reflects the neural processing of complex perceptual information. Based on this, one possible interpretation of the present data is that the N170-generating neurons influence the response, perhaps by increasing the evidence in favor of a face on the current trial. The idea that activity in a perceptual region can influence the response to an ambiguous stimulus is consistent with the mechanisms involved in binocular rivalry and motion detection. For example, Blake and Logothetis (2002) suggest that periods of left- and right-eye dominance are governed by a stochastic process with an unstable time constant. A similar principle may be at work here, such that internal stochastic activity combines with neural activity generated by the stimulus to yield the percept. While this stochastic activity may be quite weak, it may be sufficient to bias the response in favor of one alternative or the other in the absence of bottom-up or top-down evidence.

It should be noted that in studies relating the perception of motion in an ambiguous stimulus to biases in MT show that the biases appear around 50-100 ms (Britten et al., 1993), so while our effects occur early in processing, the differences at 170 ms may result from feedback from areas involved in higher-level tasks. Such top-down

effects have been noted with imagery and voluntary attention (Wojciulik, Kanwisher, & Driver, 1998; O'Craven & Kanwisher, 2000). However, the mixed design of Experiment 2 provides no incentive for feedback that might begin prior to the trial. In addition, we found no differences in the frontal electrodes nor at earlier time intervals prior to the N170, and so this explanation seems unlikely. It is possible that recurrent feedback from higher cortical areas might occur within a single trial (Hochstein & Ahissar, 2002), in which these later areas categorize the stimulus and then re-tune the face-processing areas. This account is not inconsistent with the stochastic activity account, but places the locus of the stochastic activity in higher cortical areas. If this view is accurate, it would require a reinterpretation of the N170 as reflecting not only perceptual information, but also feedback from other cortical areas that enable rapid re-tuning of the response properties of the N170 neurons within a trial. While we currently have no evidence in favor of or against this explanation, it remains an interesting interpretation.

The use of an ambiguous noise field allows us to establish a link between activity in perceptual regions and the reported percept in the absence of physical differences in the stimuli. By controlling the physical stimulus as much as possible, this technique avoids many of the stimulus differences present in experiments that initially described the response properties of the N170. This methodology can be generalized to other comparisons to reveal how ongoing neural activity affects perception and delineate how components of the ERP reflect aspects of processing on the pathway from sensation to response.

References

Bass, J. M. P., Kenemans, J. L., & Mangun, G. R. (2002). Selective attention to spatial

frequency: an ERP and source localization analysis. *Clinical Neurophysiology,*

113, 1840-1854.

Bentin, S. (1998). Separate modules for face perception and face recognition:

Electrophysiological evidence. *Journal of Psychophysiology, 12*(1), 81-81.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological

studies of face perception in humans. *Journal of Cognitive Neuroscience, 8*, 551-

565.

Bentin, S., & Golland, Y. (2002). Meaningful processing of meaningless stimuli: The

influence of perceptual experience on early visual processing of faces. *Cognition,*

86, B1-B14.

Bentin, S., Sagiv, N., Mecklinger, A., Frederici, A., & von Cramon, Y. D. (2002).

Priming visual face-processing mechanisms: Electrophysiological evidence.

Psychological Science, 13, 190-193.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996).

A relationship between behavioral choice and the visual responses of neurons in

macaque MT. *Visual Neuroscience, 13*, 87-100.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of

neurons in macaque MT to stochastic motion signals. *Visual Neuroscience, 10*,

1157-1169.

Cauquil, A. S., Edmonds, G. E., & Taylor, M. J. (2000). Is the face-sensitive N170 the

   only ERP not affected by selective attention? *Cognitive Neuroscience, 11*, 2167-

   2171.

Farah, M. (1990). Visual agnosia: Disorders of object recognition and what they tell us

   about normal vision. Issues in the biology of language and cognition. Cambridge,

   MA, US: The MIT Press.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999).

   Activation of the middle fusiform face area increases with expertise in

   recognizing novel objects. *Nature Neuroscience, 6*, 568-573.

Goffaux, V., Gauthier, I., & Rossion, B. (2003). Spatial scale contributions to early visual

   differences between face and object processing. *Cognitive Brain Research, 16*,

   416-424.

Hochstein, S. & Ahissar, M. (2002). View from the top: Hierarchies and reverse

   heirarchies in the visual system. *Neuron, 36*, 791-804.

Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp.

   Experimental Brain Research, 78, 193-202.

Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for

   faces not animals. *NeuroReport, 10*, 183-187.

Martinez, A., Di Russo, F., Anllo-Vento, L., & Hillyard, S. A. (2001).

   Electrophysiological analysis of cortical mechanisms of selective attention to high

   and low spatial frequencies. *Clinical Neurophysiology, 112*, 1980-1998.

McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of

    human face perception II: Response properties of face-specific potentials

    generated in occipitotemporal cortex. *Cerebral Cortex, 9*, 431-444.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates

    corresponding stimulus-specific brain regions. *Journal of Cognitive*

    Neuroscience, 12, 1013-1023.

Olivares, E. I., & Iglesias, J. (2000). Neural bases of perception and recognition of faces.

    Revista De Neurologia, 30(10), 946-952.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition.

    Psychological Science, 12, 43-47.

Tong, F., Nakayama, K., Vaughan, J.T., &Kanwisher, N. (1998). Binocular rivalry and

    visual awareness in human extrastriate cortex. *Neuron, 21,* 753-759.

Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates

    face-specific activity in the human fusiform gyrus: fMRI study. *Journal of*

    Neurophysiology, 79, 1574-1578.

Effect Size and Probability of Responding "face"	
P("face")	**Effect Size (μV)**
0.11	2.78
0.29	7.60
0.43	6.21
0.47	2.71
0.60	-2.06
0.67	6.20
0.72	0.90
0.93	17.83
0.96	33.12
0.97	15.38

Table 1. Probability of responding "face" to a noise-alone stimulus and effect size data for all observers. Effect size is computed by subtracting the N170 amplitude to face responses from the N170 amplitude to word responses. Positive numbers imply a greater N170 when observers report seeing a face.

Average Median Reaction Times		
(in milliseconds)		
	Response	
Stimulus	**Said 'Face'**	**Said 'Word'**
Noise Alone	615 (20)	615 (23)
	Face Presented	**Word Presented**
Low Contrast	508 (15)	569 (17)
High Contrast	488 (14)	506 (13)

Table 2. Median reaction times averaged across subjects for Experiment 2. The averages and standard errors are shown for noise-alone stimuli binned by response, and for correct responses to high and low contrast faces and words.
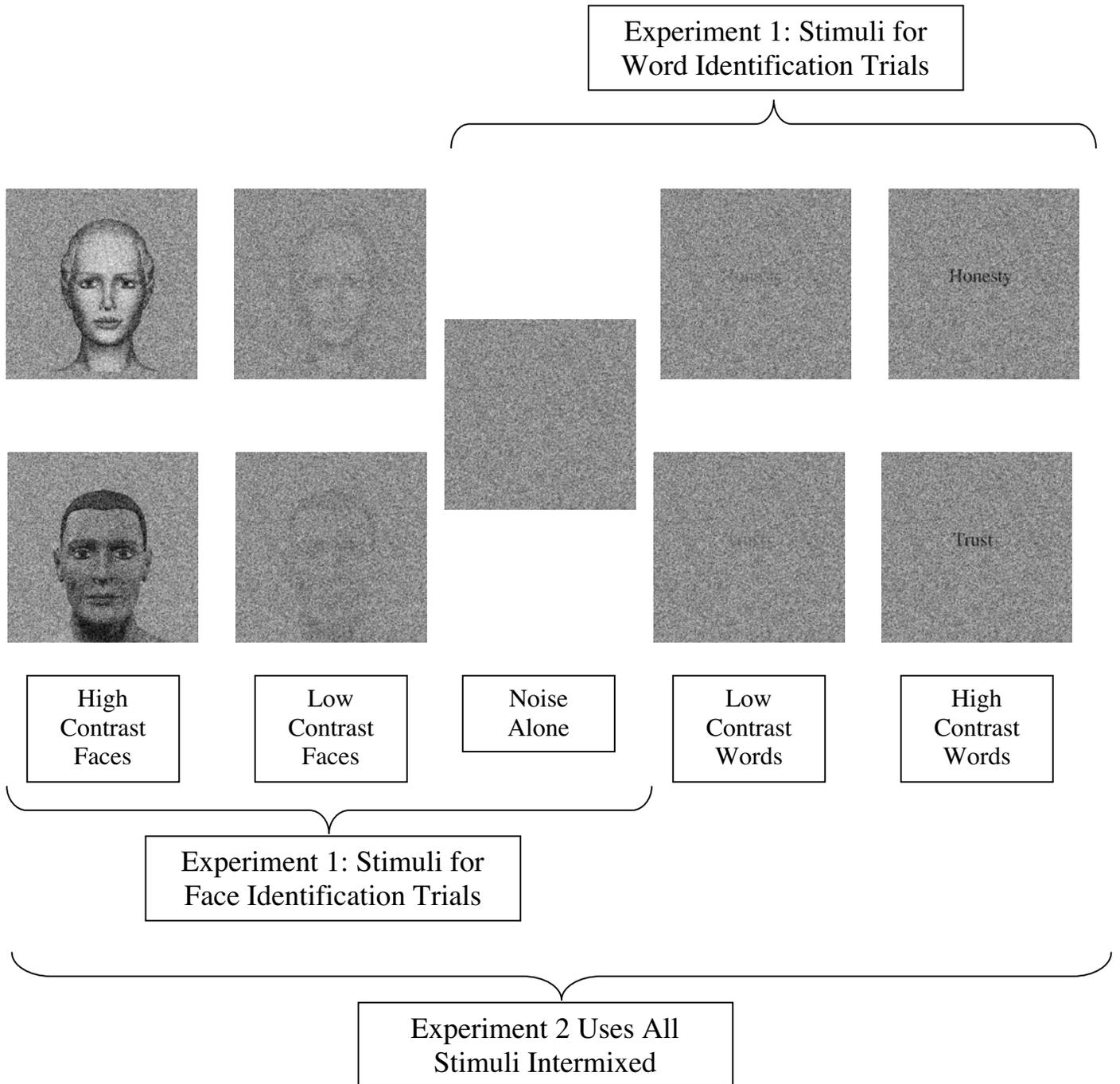
Figure 1. All stimuli used in both Experiments. Experiment 1 presented the faces and words in a blocked design, with the noise-alone stimulus appearing in both blocks. The Experiment 2 design presented all stimuli intermixed across trials. Note that the noise is identical for all stimuli, and the figures above represent the complete stimulus set.
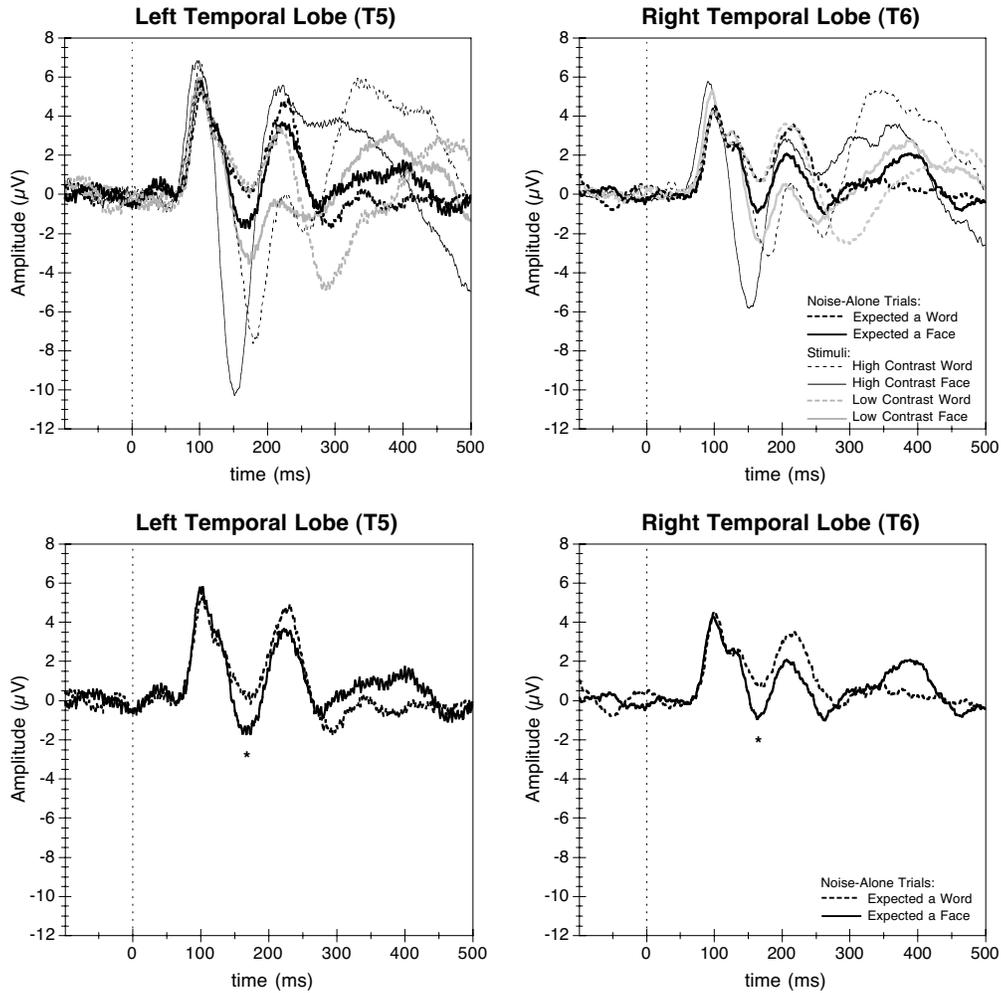
Figure 2. Data from Experiment 1. ERPs elicited at temporal lobe sites T5 (left panel, left hemisphere) and T6 (right panel, right hemisphere). All data are shown in the top panel; darker lines indicate noise-alone trials. These data are shown alone in the bottom panel. The asterisks in the left and right panels indicate significant differences between the two responses at the N170 component in both channels. The N170 component is the large dip that occurs between 100 and 200 ms.
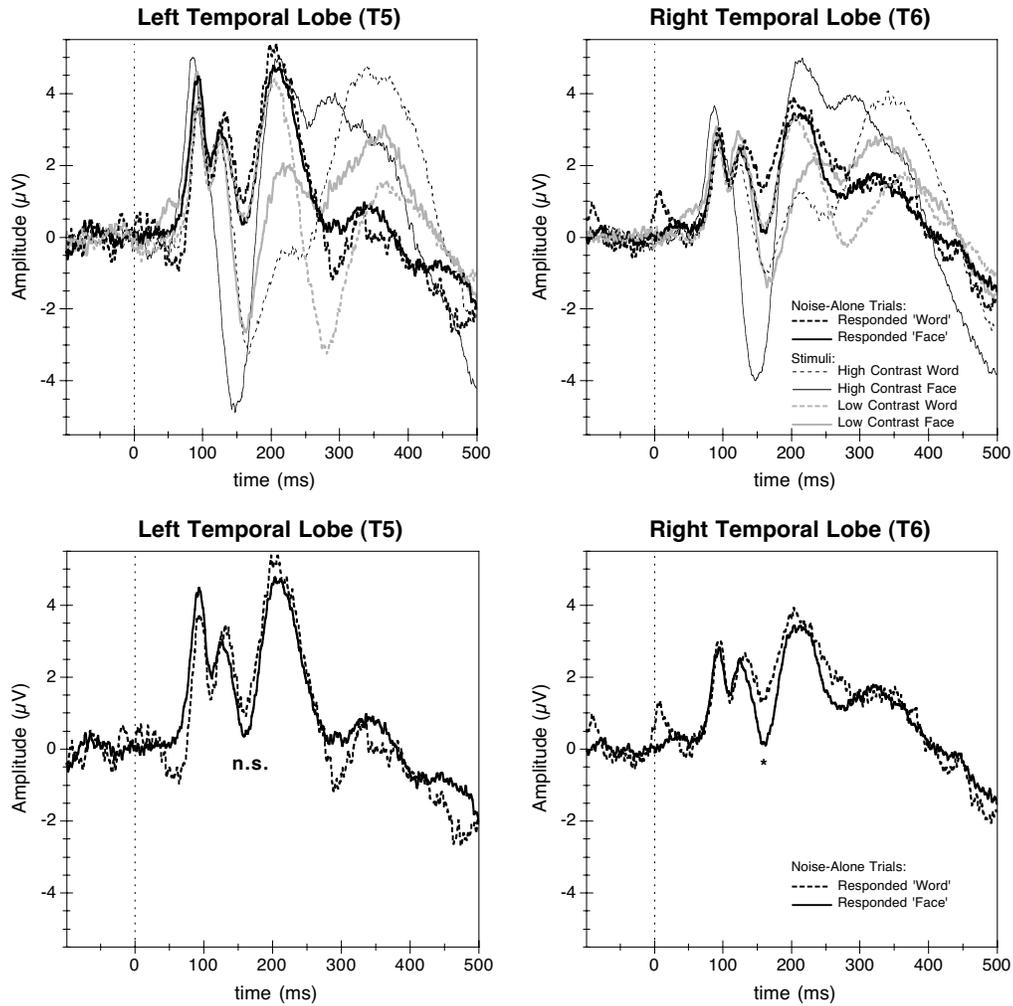
Figure 3. Data from Experiment 2. ERPs elicited at temporal lobe sites T5 (left panel, left hemisphere) and T6 (right panel, right hemisphere). All data are shown in the top panel; darker lines indicate noise-alone trials. These data are shown alone in the bottom panel. The N170 component is the large dip that occurs between 100 and 200 ms. The asterisk in the right panel indicates significant differences between the two responses at the N170 component. The abbreviation n.s. indicates no significant difference.
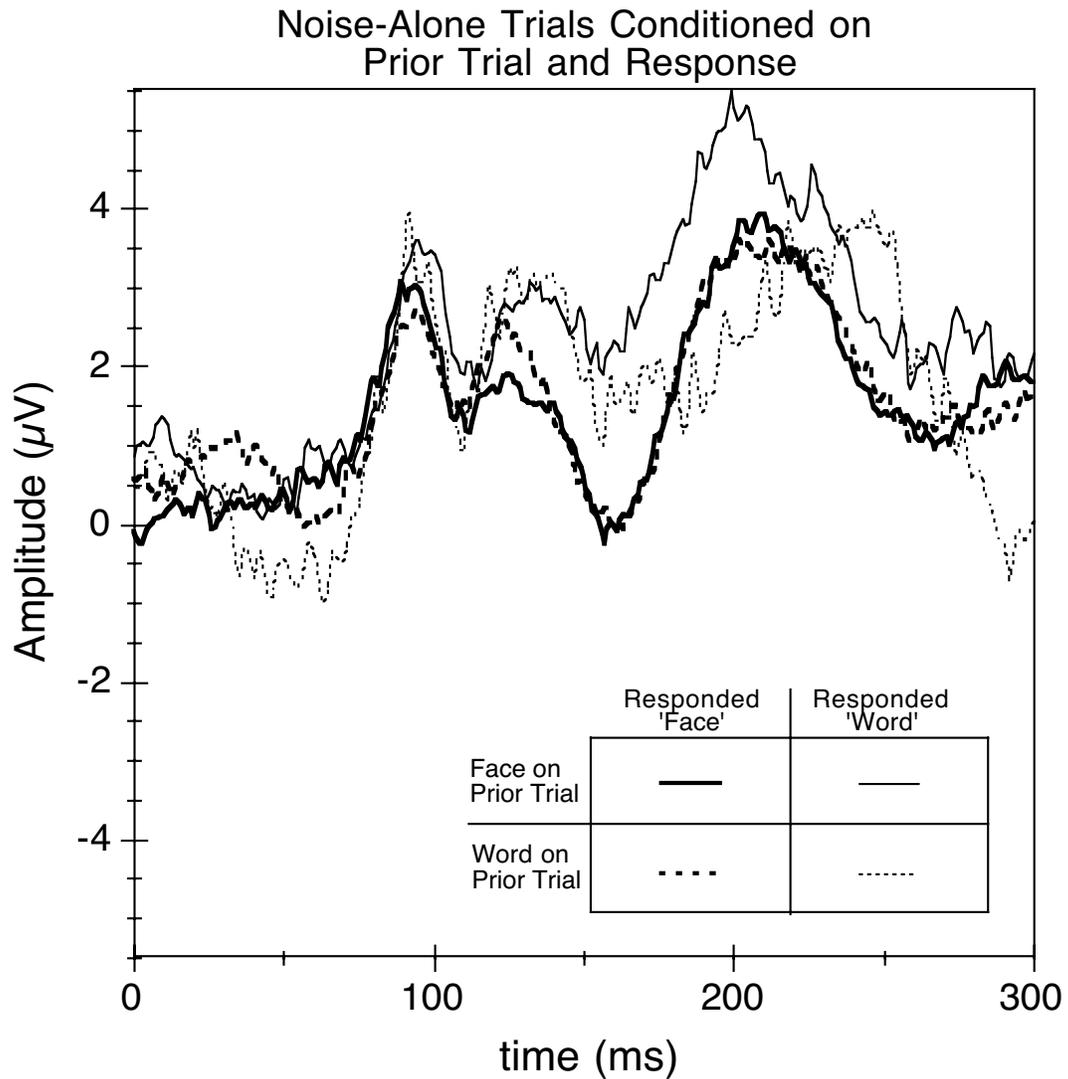
Figure 4. Data from Experiment 2, channel T6 (right hemisphere). ERP traces from noise-alone trials conditioned on prior trial and response. Note that the time scale is different than in prior figures to emphasize effects at the N170.