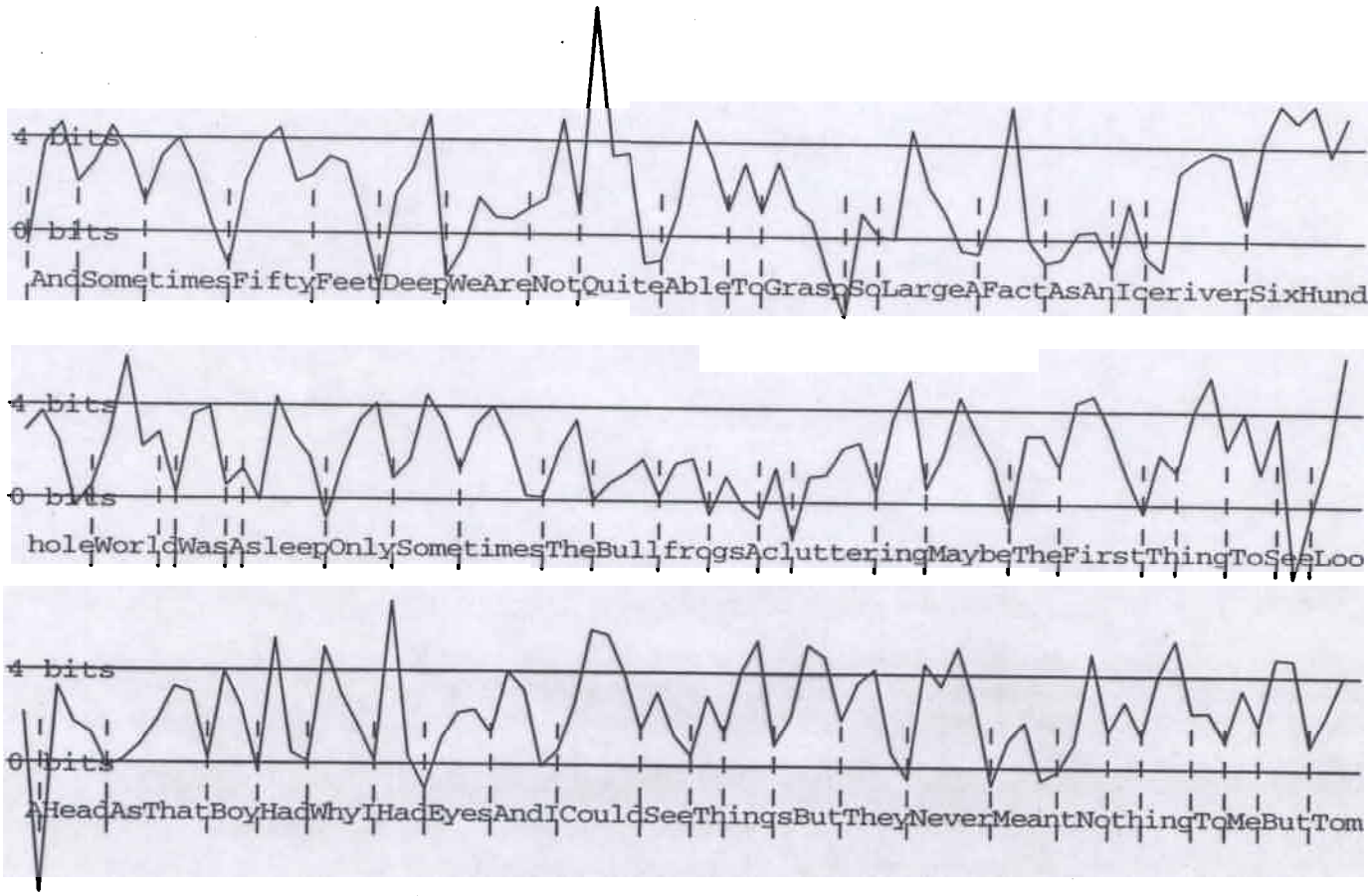


Figure 1.1: Binding energy of 4-tuples in three selections from Mark Twain.

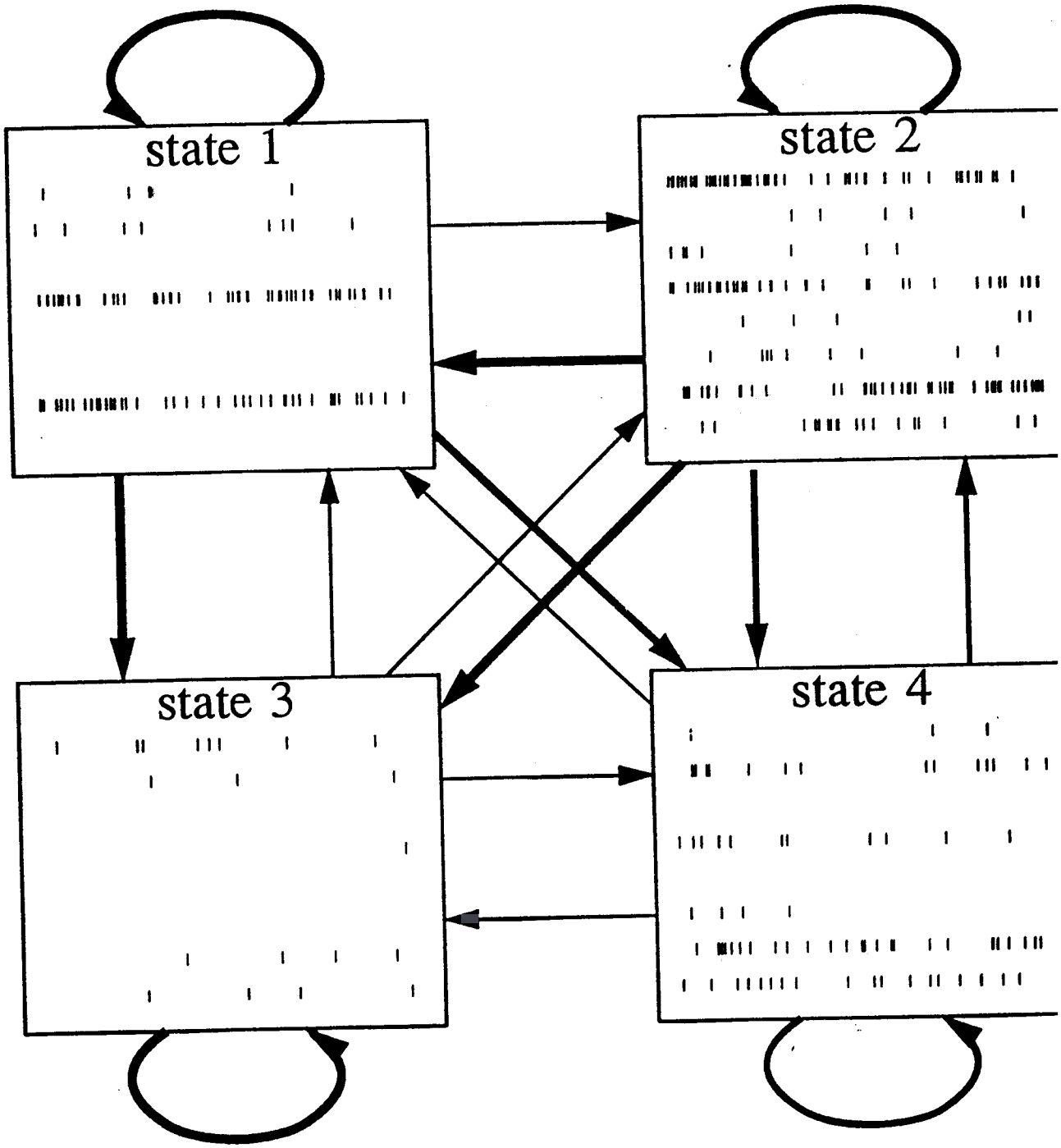


Fig 1

Figure 4: A set of texture images and the output of the boundary system in response to these images. Parameters used for all but 'Sierra.im' are $\lambda = 6, \alpha = 0.02, \gamma_a = 2, \gamma_\theta = 1$. For 'Sierra.im', $\lambda = 12, \alpha = 0.005, \gamma_a = 4, \gamma_\theta = 4$.

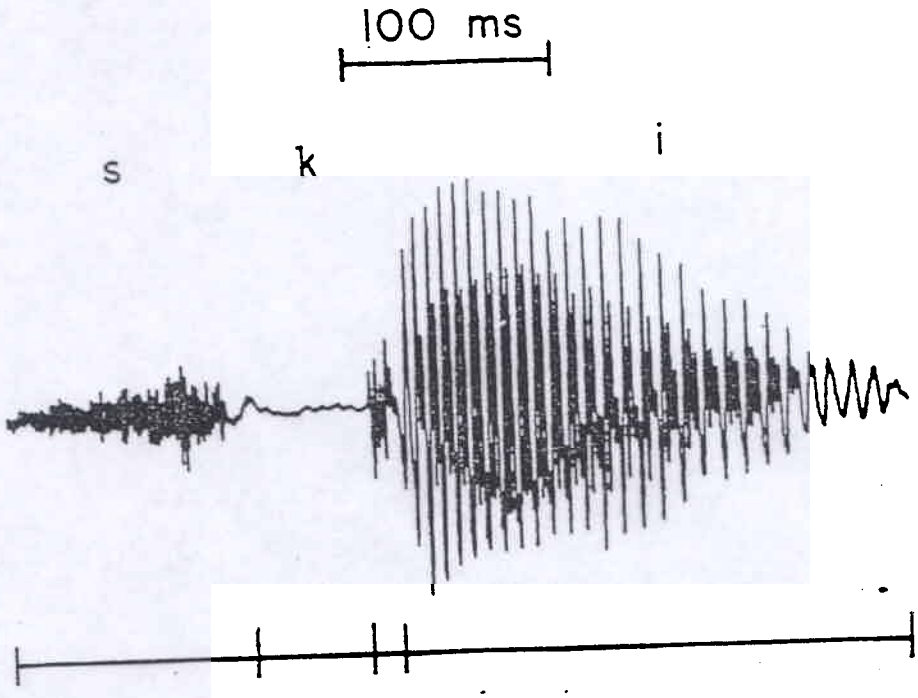


Figure 1: Acoustic waveform for an utterance of the word *SKI*.

Step II: Boundaries between simple units

- Text: 'binding energy' between letters

  For all 4-tuples, compute $p_4(abcd)/\left(p_2(ab)\cdot p_2(cd)\right)$

- Spike trains as Markov chains between states

  Observing n neurons, let their rates $(r_1,\ldots,r_n)$ jump from state to state.

- Speech as Markov chains of phonemes

  Let the base frequency f and power spectrum $\{a_n\}$ jump between states

  esp. $f=\infty$ (unvoiced) and $f=f_0$ (voiced)

- Images as texture patches

  Let $\{E_\alpha, F_\alpha\}$ change across subregions of the image domain: boundaries are now curves and need their own stochastic model

Step III: Domain Warping

- Spike trains: continuously varying spike rate $r(t)$

- Speech: variable frequency formants (peaks in power spectrum) and variable length phonemes

- Images: e.g. textures are warped by varying scale and orientation; faces are warped by changing viewing angle, expression, proportions

WHY IS THIS A MAJOR COMPLICATION?

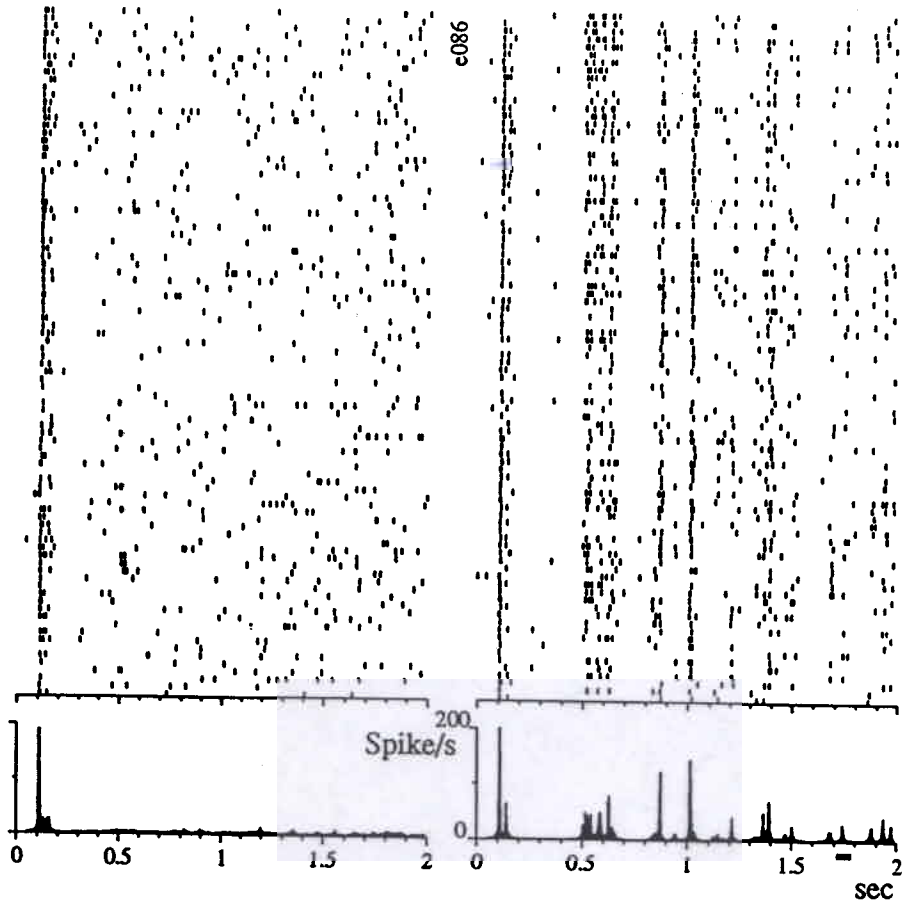


Figure 1 The neuronal response of one cell in area MT in a fixating macaque monkey to randomly seeded dynamic dot stimuli at $c = 0$ presented for 2 sec (left) is grossly characterized, following the initial transient, by a random point process with a mean rate of 3.4 Hz. However, when a dynamic dot stimulus formed with a *particular* seed is repeated (but interleaved with different stimuli) many times, the reliability of the response becomes apparent (right). Viewing the spike trains from an acute angle reveals the precision of the pattern, which is quantified in Fig. 2.

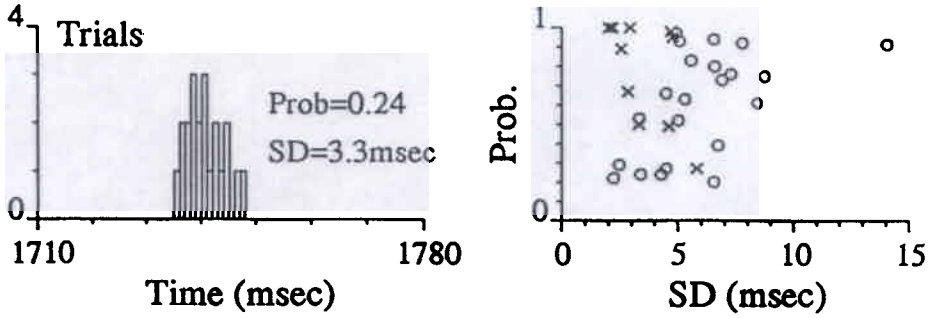
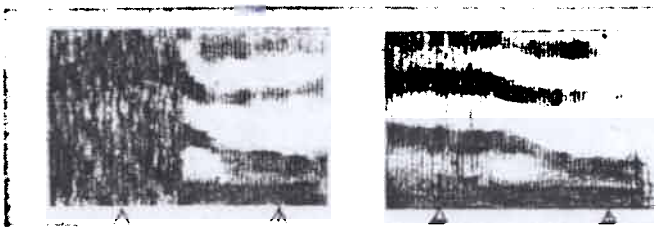
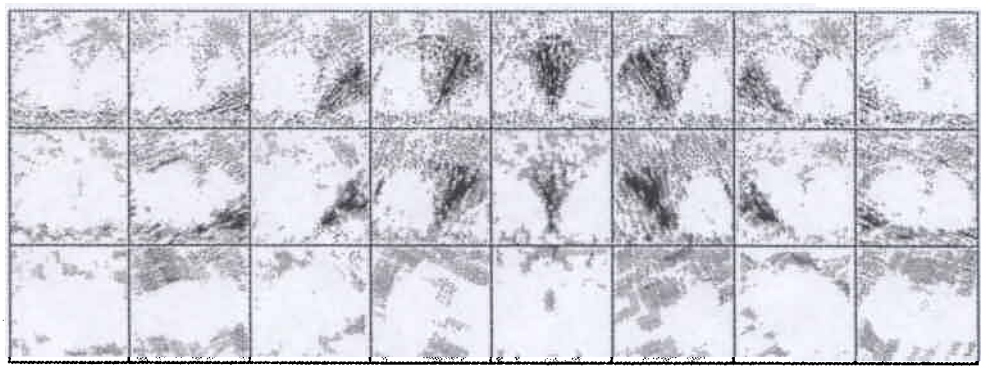


Figure 2 To estimate the precision of the neuronal response, we measure the standard deviation (SD) in time of responses that are clearly isolated from other action potentials. For example, the distribution of the start time of the response at 1740 msec in Fig. 1 (bar, bottom right) is shown here at the left. For the 24% of trials which have action potentials within the period 1710–1780 msec, the distribution of the first spike in the response has SD=3.3 msec. Using this method, the estimated maximum precision (minimum SD) is plotted against the probability of response for 22 of 26 cells from monkey E (right, circles). Four cells were not suited for this analysis due to high background firing rates or lack of modulation. Crosses indicate initial transients (not always present) for the same set of cells.

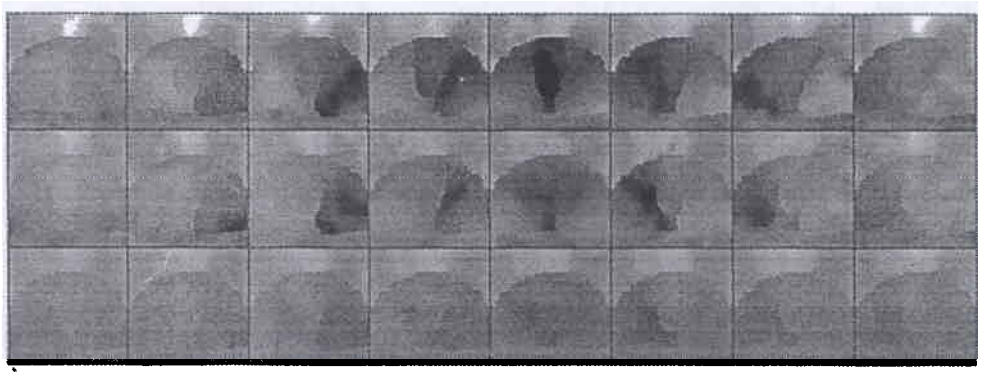


sue *zoo*

si (*see*) sɛd (*said*) su (*sue*)



Initial response map of the texture surface channels.



Final response map of the texture surface channels.

Figure 5: Surface responses to 'Shell.im'. The responses for each map have been normalized with respect to the largest response in the map. Within each map, there are 3x8 little squares. Each square shows the spatial response of a particular channel. First column on the left is horizontally tuned in orientation. Orientation rotates with column. Top row is the finest scale. Scales increase with row. The initial response is equal to direct input from simple cells, and is very localized in space. The final response shows significant diffusion in all dimensions, with the shell's unifed surface cartoon appearing in every channel.

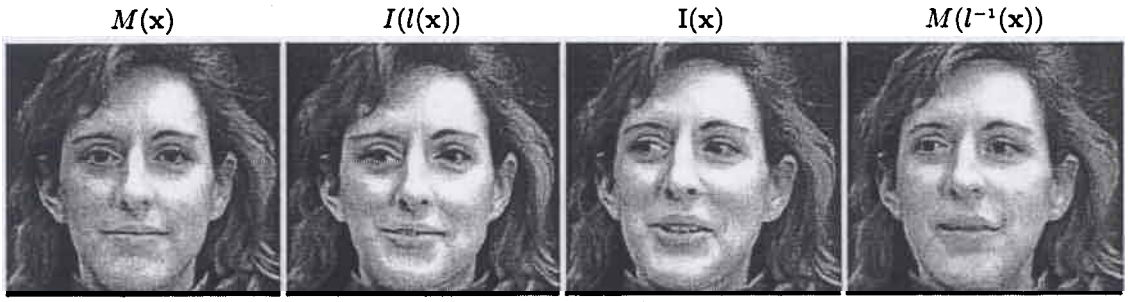


Figure 4.16: Example demonstrating that a local warp can successfully capture small changes in viewpoint. The eyes change their direction of gaze only slightly, because moving more would entail introducing portions of the whites that are not present in the input image (or conversely deleting whites from the model image).

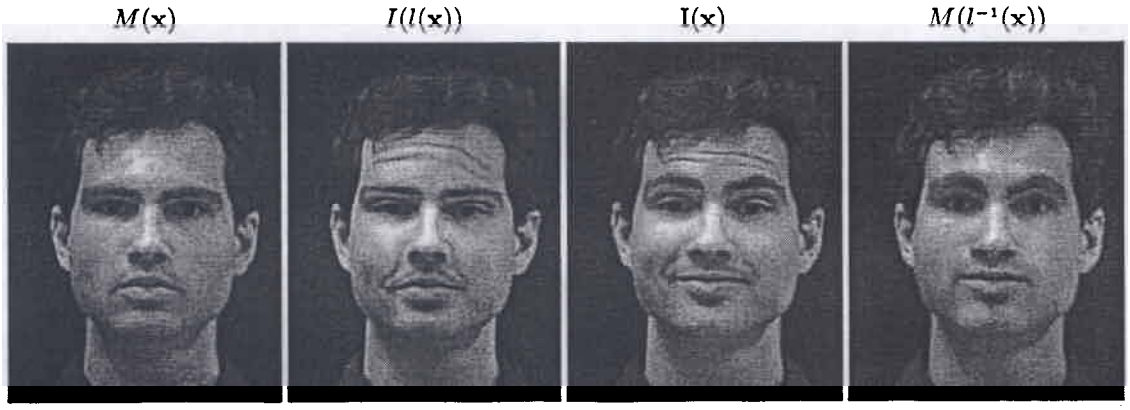


Figure 4.17: Example demonstrating what happens when there is a change of expression. The wrinkles induced by smiling have no match in the model and so can at best be shrunk into thinner lines or pushed into slightly more shaded areas. Dimples are similarly impossible to remove. The warping can easily distort wrinkles and dimples enough so that residue analyses may fail unless they model the errors induced by the face model.