

# Spatial-frequency bands in complex visual stimuli: American Sign Language

Thomas R. Riedl\* and George Sperling

Human Information Processing Laboratory, Department of Psychology, New York University, New York,  
New York 10003

Received April 27, 1987; accepted December 10, 1987

Dynamic images of individual signs of American Sign Language (ASL) with a resolution of  $96 \times 64$  pixels were bandpass filtered in adjacent frequency bands. Intelligibility was determined by testing deaf subjects fluent in ASL. The following results were obtained. (1) By iteratively varying the center frequencies and bandwidths of the spatial bandpass filters, it was possible to divide the original signal into four different component bands of high intelligibility. (2) The measured temporal-frequency spectrum was approximately the same in all bands. (3) The masking of signals in band  $i$  by noise in band  $j$  was found to be inversely proportional to  $\log |f_{\text{signal}}/f_{\text{noise}}|$ . At constant performance, the ratio of root-mean-square signal amplitude to noise amplitude,  $s/n$ , was the same for bands 2, 3, and 4 and higher for band 1. (4) When weak signals  $i$  and  $j$  were added linearly, there was a slight intelligibility advantage for signals in the same band ( $i = j$ ) compared with signals in adjacent bands and for signals in adjacent bands compared with signals in distant bands.

## INTRODUCTION

Much has been learned about how the spatial-frequency components of simple visual stimuli, in combination, contribute to visual responses. Most of what we know is concerned with simple stimuli near their threshold.<sup>1</sup> For example, there is ample evidence that multiple channels (mechanisms) are involved in the detection of simple visual stimuli—different channels at different retinal spatial frequencies.<sup>2</sup> It is believed that, at threshold, these channels sum their information probabilistically. Whether a channel that subserves one spatial frequency inhibits channels that subserve other frequencies is unclear; different results are reported for different procedures.<sup>1</sup>

Much of the visual research is concerned with spatial frequencies as they are produced at the retina. The discriminability of stimuli that are well above threshold, and explicitly limited by external noise, is independent of viewing distance (retinal angle) over a wide range.<sup>3,4</sup> Noisy signals are discriminated equally at vastly different retinal frequencies, and their perceptual properties are best characterized by cycles per object rather than cycles per degree of visual angle.

In a visual communication channel for complex, dynamic visual stimuli, such as American Sign Language (ASL), the limitations are related to stimulus noise and to stimulus subsampling rather than to low contrast; that is, the intelligibility of these ASL stimuli is limited by external distortions, modeled as noise, rather than by internal noise. Such limitations will probably be characterized by object spatial frequencies,<sup>5</sup> and almost none of the previous literature on spatial-frequency interactions in vision is directly applicable. Therefore, to design optimal communication channels for transmitting dynamic complex stimuli, there is no alternative to studying them directly.

From a practical point of view, visual communication channels would be immediately useful to the several hundred thousand hearing-impaired individuals who rely on

ASL for communication.<sup>6</sup> More than two million Americans are unable to understand speech even with a hearing aid; many of these would benefit by having a visual communication channel to aid their utilization of residual hearing. The problem is that available, affordable channel capacity is limited, and compressing images to utilize this capacity efficiently requires a better understanding of how frequency components of complex images contribute to their intelligibility as well as better methods of image compression.<sup>7-9</sup>

This study is concerned with how the visual information in component spatial-frequency bands of a complex visual signal, ASL, combines to facilitate or to interfere with the intelligibility of ASL. Therefore first we attempt to establish four spatial-frequency bands having approximately equal intelligibility for ASL. Second, we measure the temporal characteristics of each of these bands. Third, we study how various intensities of noise in frequency band  $i$  interfere with signals in band  $j$ . Fourth, we determine how weak signals in band  $i$  combine with weak signals in band  $j$  to facilitate perception.

## EXPERIMENT 1: BANDS OF EQUAL INTELLIGIBILITY

The purpose of experiment 1 is to derive a number of spatial-frequency filters to produce bandpass ASL stimuli from the original ASL stimuli. Each band should have approximately equal, and moderately high, intelligibility. Preliminary work suggested that four such bands would be possible for our stimuli.

### Method

#### Original Stimuli

The stimuli consisted of isolated ASL signs displayed at 30 frames per second (fps) on a television raster monitor. Signs took 2–3 sec and consisted of 60–90 frames. A stan-

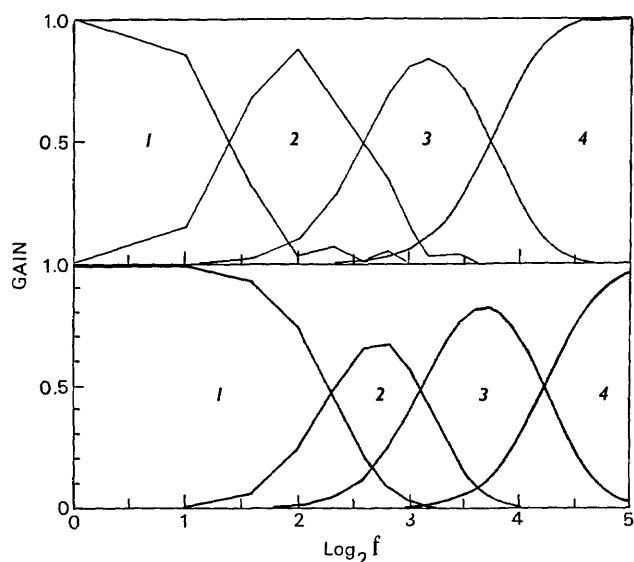


Fig. 1. Gain versus frequency for the filters used to create the spatial bandpass stimuli. Frequency is in  $\log_2$  (cycles per frame width). Upper graphs represent the filters used for the initial investigation (experiment 1a); lower graphs represent the filters used in experiment 1b and all subsequent experiments. The numbers 1-4 are used to designate the filter bands.

standard starting and ending position of the signer was used for all signs. A sign was initially photographed on 16-mm movie film (at 30 fps) and digitized to  $512 \times 512$  pixels per frame. It was then reduced and cropped to  $96 \times 64$  pixels and embedded in a uniform background that extended to  $128 \times 128$  pixels. The background luminance equaled the mean luminance of the ASL sequence. Pixel intensity was represented by 256 discrete gray levels. These are small, low-resolution images, but they are essentially as intelligible as a natural, direct view of the signer. Mistakes occur primarily when the viewer is unfamiliar with a sign.<sup>9</sup> For more details of the stimuli and procedures, see Refs. 9 and 10.

#### Bandpass Filters

The filters used to produce the bandpass stimuli should have the following characteristics:

(1) They should represent adjacent frequency bands with as little overlap as possible. However, the steepness of

the filter cutoffs in the frequency domain must be balanced against the ripple in the spatial domain caused by steep cutoffs.

(2) There should be no spatial-phase distortion.

(3) The sum of signals in all the filters should exactly equal the unfiltered original (reconstructable from components).

(4) The boundaries between bands should be continuously adjustable.

(5) The bands should be of approximately equal and of moderately high intelligibility.

The desired filter characteristics were obtained by generating the filters iteratively from Gaussian functions. Pyramids based on differences of Gaussians are well known.<sup>11</sup> The filter scheme used here is not formally a pyramid because all bands are represented by the same number of spatial samples, but it is quite similar. Here, we use differences based on iterated Gaussians to produce filters whose center position and bandwidth in frequency space can both be varied. This enables us to divide frequency space ( $\omega_x, \omega_y$ ) into an arbitrary number of slightly overlapping, concentric annular regions (the filters) whose boundaries are adjustable. The summed output of all filters equals the original signal. The filter-generating algorithm is described in detail in Appendix A. The initial instantiation of the algorithm, filter set 1 for experiment 1, is shown in Fig. 1.

#### Experiment 1a: Filter Set 1

##### Procedure

The stimuli for experiment 1 consisted of 100 isolated ASL signs divided randomly into four groups and filtered in the four spatial-frequency bands of filter set 1. The processed stimuli were recorded on Betamax I video-recording cassettes and displayed on a television monitor where the  $128 \times 128$  display of the video-recorded display subtended  $5.3 \times 8.4$  cm (horizontal  $\times$  vertical). The screen was viewed through a viewing hood from a distance of 56 cm. Each sign was preceded by a visual warning cue presented 2 sec before the sign. Subjects were instructed to write an English word for the ASL sign on a prepared answer sheet and to make their best guess when they were uncertain. Four deaf subjects were recruited through New York University and local organizations for the deaf. They were instructed in the

Table 1. Filter Parameters in Cycles per Frame Width<sup>a</sup> and the Measured Intelligibility of the Filtered ASL Signs

Experiment and Filter Set	Band	Frequency (Cycles per Frame Width)					% Correct
		Low Half-Power	Peak	High Half-Power	2D Mean Amplitude <sup>b</sup>	2D Mean Power <sup>b</sup>	
Experiment 1a, set 1	—	—	2.3	4.2	1.6	38.0	
	2	3.0	4.0	4.9	6.1	4.5	60.0
	3	7.1	9.0	10.6	10.7	9.9	67.7
	4	15.3	—	—	22.7	23.8	80.3
Experiment 1b, set 3	1	—	—	4.2	3.0	2.5	66.4
	2	4.8	6.5	8.6	7.5	7.0	67.6
	3	9.3	12.5	17.6	15.2	14.1	87.5
	4	21.5	—	—	24.7	26.2	80.1

<sup>a</sup> To obtain the frequency in cycles per centimeter (in the object domain), divide the frequency by 30.5 cm per frame width (the field width at the signer's head). Four subjects were used in experiment 1; eight were used in experiment 3.

<sup>b</sup> The frequency components used to compute mean frequencies do not include  $f = 0$ . 2D, two-dimensional.

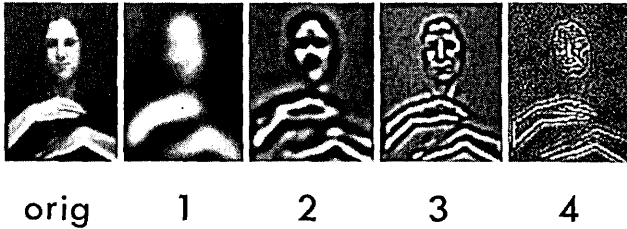


Fig. 2. The ASL images filtered in bands 1-4. The leftmost image is the unfiltered original.

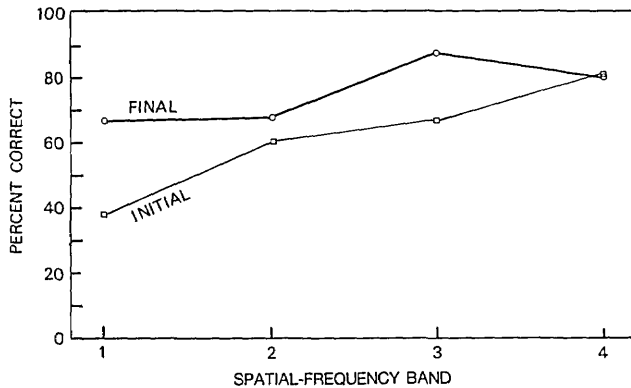


Fig. 3. Intelligibility (percentage of correct ASL sign identifications) as a function of the spatial-frequency band. Curve labeled INITIAL was obtained in experiment 1a with the filter set at top of Fig. 1; curve labeled FINAL was obtained in experiment 1b with filters at the bottom of Fig. 1 and with improved stimuli.

procedure by a proficient signer. The signs were run in blocks (by frequency band) so that the signer would be maximally prepared for the type of stimulus to be shown on a trial.

### Results

The average percentages of correct responses in each band are shown in Table 1. As can be seen, performance improves with increasing frequency, from 38% in band 1 to 80% in band 4.

### Experiment 1b: Filter Set 3

#### Procedure

Filter set 1 did not generate equally intelligible bands. Therefore the filters were changed according to an algorithm that estimated the contribution to intelligibility of every component frequency and attempted to distribute these contributions equally among the bands. In addition to intelligibility differences among bands in experiment 1a, we noted that there were some unfamiliar signs and that these may not have been distributed equally among groups. Therefore, for subsequent tests, 28 ambiguous signs were discarded. The remaining 72 signs were divided into four groups and were tested as before. Subsequently, the filters were again adjusted by an algorithm to increase the bandwidth of the bands with the worst performance and to diminish the bandwidth of the bands with the best performance. The final filters are shown in Fig. 1, and examples of the filtered stimuli are illustrated in Fig. 2.

To make the intelligibility test more accurate, data collected up to this point were used to rank the signs into three

categories: easy, medium, and difficult. Signs in each category were distributed evenly into band conditions. Further, a balanced Latin square block design was used so that each sign was processed in each frequency band; i.e., four complete stimulus video tapes were prepared, each of which contained all the experimental ASL signs but distributed into different filter groups. Eight subjects were run, two subjects for each cell of the Latin square.

### Results

Filter set 3 yielded four bands with intelligibilities that were more nearly equal than those of filter set 1, but intelligibility was still not completely uniform across bands. Intelligibility ranged from 66% in band 1 to 87% in band 3 (Fig. 3). Although the four bands of filter set 3 were not equally intelligible, they were sufficiently close to equal that we could move forward with the main experiments to investigate how signals in different bands interfere with and facilitate one another.

### EXPERIMENT 2: THE TEMPORAL-FREQUENCY SPECTRUM

Here we address the question: What is the temporal power spectrum of the signal in each of the spatial bands derived in experiment 1? This question is of interest in its own right in terms of discovering the correlation of spatial and temporal frequencies in the environment and therefore in defining the optimal visual detectors for operating in this environment. More immediately, we will need the temporal data in experiment 3 to create dynamic visual noise that is matched to the spatially band-limited ASL signals in both spatial and temporal frequency.

To determine the signal power as a function of temporal frequency, eight representative ASL signs were selected. At the mean spatial frequency  $m_i$  of each spatial-filter  $i$  of experiment 1 (see Table 1, column 6), a small spatial-frequency range  $\Delta m_i$  [ $\Delta m_i = \{(\omega_x, \omega_y) | m^2 - \epsilon \leq \omega_x^2 + \omega_y^2 < m^2 + \epsilon\}$ ] was selected for analysis. This is the range of spatial frequencies that best characterizes its spatial-frequency band.

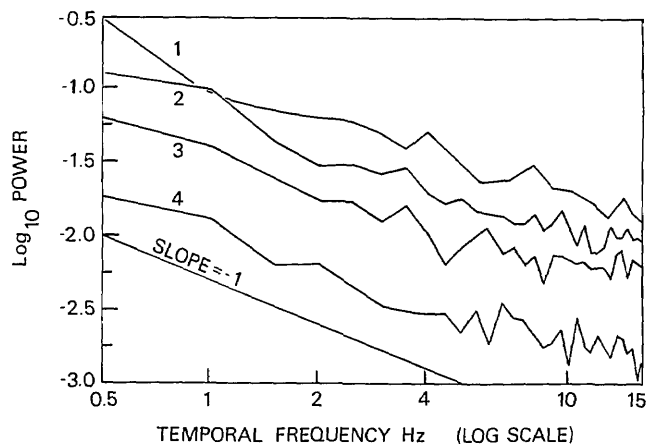


Fig. 4. The temporal power spectrum of ASL in spatial-frequency bands 1-4. The abscissa represents the temporal frequency in hertz; the maximum frequency of 15 Hz is determined by the frame rate of 30 Hz. The ordinate represents the average power in an annular band of temporal frequencies extracted from a three-dimensional  $(x, y, t)$  Fourier analysis of eight representative ASL sign sequences. The line of slope  $-1$  is drawn for reference.

The spatial range  $\Delta m_i$  is an annulus in  $\omega_x$ - $\omega_y$  spatial-frequency space and a hollow cylinder in  $(\omega_x, \omega_y, \omega_t)$  spatio-temporal-frequency space. For every small range of temporal frequencies  $\Delta f$  within  $\Delta m_i$  the average power (over the eight signs) was computed at each spatiotemporal frequency (annular cross section of the cylinder). The whole computation was repeated for each of four spatial bands  $i$ . These data (temporal power versus temporal frequency, for each of the four spatial frequencies  $m_i$ ) are displayed in Fig. 4.

Overall temporal power diminishes with increasing spatial frequency. Within each spatial-frequency band, temporal power falls off with an initial slope of approximately  $-1$  on the graph of  $\log_{10}$  (power) versus  $\log_{10}$  (frequency), leveling off at high temporal frequencies. The approximate parallelism of the temporal-frequency power curves (for different spatial frequencies) suggests that the temporal-frequency composition of our ASL stimuli is independent of their spatial composition.

### EXPERIMENT 3: CROSS-BAND MASKING BY NOISE

Typically, cross-band masking has been studied with simple static signals<sup>1,2,12-18</sup> rather than with realistic dynamic stimuli. The purpose of experiment 3 is to determine the extent to which dynamic noise in spatial-frequency band  $j$  interferes with dynamic ASL signals in band  $i$ , for all 16 combinations of  $i, j, = 1, 2, 3, 4$ . Basically, this requires determining the performance versus the signal-to-noise ratio in each of the 16 different band combinations. Because at least half a dozen values of  $s/n$  must be sampled to determine a performance function, this experiment requires determination of the performance in almost 100 conditions. Since it is impractical to create and maintain a stimulus set of ASL signs large enough for this immense task, a rating procedure was used instead that involved intelligibility judgments of only two representative ASL signs.

#### Method

##### Stimuli

The signals were the recorded ASL signs "home" and "flower" from the previously described set. They were filtered in each of the four bands determined by filter set 3 of experiment 1 (Fig. 1). To generate noise stimuli, we started with white Gaussian noise in  $(x, y, t)$ . In the frequency domain, the noise power spectrum was shaped, separately in each of the four bands, to conform to the three-dimensional  $(x, y, t)$  power spectrum of the signals; that is, within each spatial-frequency band, the temporal shape of the noise power spectrum was matched to the shape of the signal temporal spectrum as determined in experiment 2.

##### Signal Power in a Frame

The signal power in a frame is defined as the variance of the signal luminance over the pixels of that frame. The signal power  $\sigma_s^2$  is the average power of the frames in a sequence. (In fact, the power variation between frames is small.) The noise power  $\sigma_n^2$  is computed similarly.

##### Signal-to-Noise Ratio

The signal-to noise ratio  $s/n$  is  $\sigma_s/\sigma_n$ . Note that here the signal-to-noise ratio is defined in terms of standard devi-

ations, the root-mean-square (rms) amplitudes of the signal and the noise. These are the square roots of the powers of the signal and the noise. A set of stimuli illustrating the noise, the signals, and their combinations is shown in Fig. 5.

##### Procedure

The display viewed by the subject consisted of two adjacent sequences. On the left-hand side was a noiseless sign in band  $i$ , and on the right-hand side the same ASL sign filtered in the same band  $i$  was combined with added noise from band  $j$ ; 176 such pairs were presented to the subjects. The combinations of  $i, j, s/n$ , and the ASL sign occurred in random order.

##### Rating Scale

Subjects viewed the noisy and noiseless sequences side by side and were asked to rate the noisy one on the following rating scale:

- 0, Cannot detect sign at all;
- 1, Barely visible sign, but cannot see sign;
- 2, Visible sign, some trace of sign;
- 3, Can guess at sign, but most features indiscriminable;
- 4, Fairly discriminable sign, but some critical features missing;
- 5, Visible sign, but poor-quality image;
- 6, Highly discriminable sign with good-quality image.

Subjects used fractional ratings to describe their judgments more precisely. The noiseless sequences served as references to help the subjects anchor their responses. Ratings were collected from three subjects. Subsequently, the  $s/n$  values were adjusted to obtain a better sample of the rating function, and three more subjects were run. In this experiment alone, the subjects were hearing nonsigners.

#### Results

The stimulus range was quite large, from stimuli in which the subtle details of an ASL sign were perfectly visible to stimuli in which even the presence of the signer was completely masked by noise. Thus the range of ratings, for any particular stimulus condition, was rather small. Within this range, it was most practical simply to treat the ratings numerically and to obtain the average rating across subjects. In a previous study,<sup>9</sup> quality ratings were obtained for a large set of stimuli, a subset of which was then carefully tested by formal intelligibility tests. The correlation between rated quality and objectively measured intelligibility was 0.85. Considering that the intelligibility-tested stimuli were a homogeneous subset of the most-intelligible stimuli, the high correlation was, in the authors' words, "an impressive vindication of the rating procedure" (Ref. 9, p. 364).

Figure 6 shows an example of 1 of the 16 rating-versus- $s/n$  functions for stimulus band 3 with noise band 3. The data (mean rating  $R$  versus  $\log s/n$ ) were fitted by three-segment linear functions (a total of three parameters) constrained as follows. ( $s$  and  $n$  are shown as  $S$  and  $N$  in all the figures.)

In segment 1, the left-hand asymptote was constrained to be horizontal at  $R = 0$ . In segment 3, the right-hand asymptote as  $s/n \rightarrow \infty$  was horizontal at  $R = R_\infty$ . Segment 2 connected segments 1 and 3. The square deviation of the data from the three-segment fit was minimized by an optimization program.<sup>19</sup> Figure 6 illustrates the parameter-

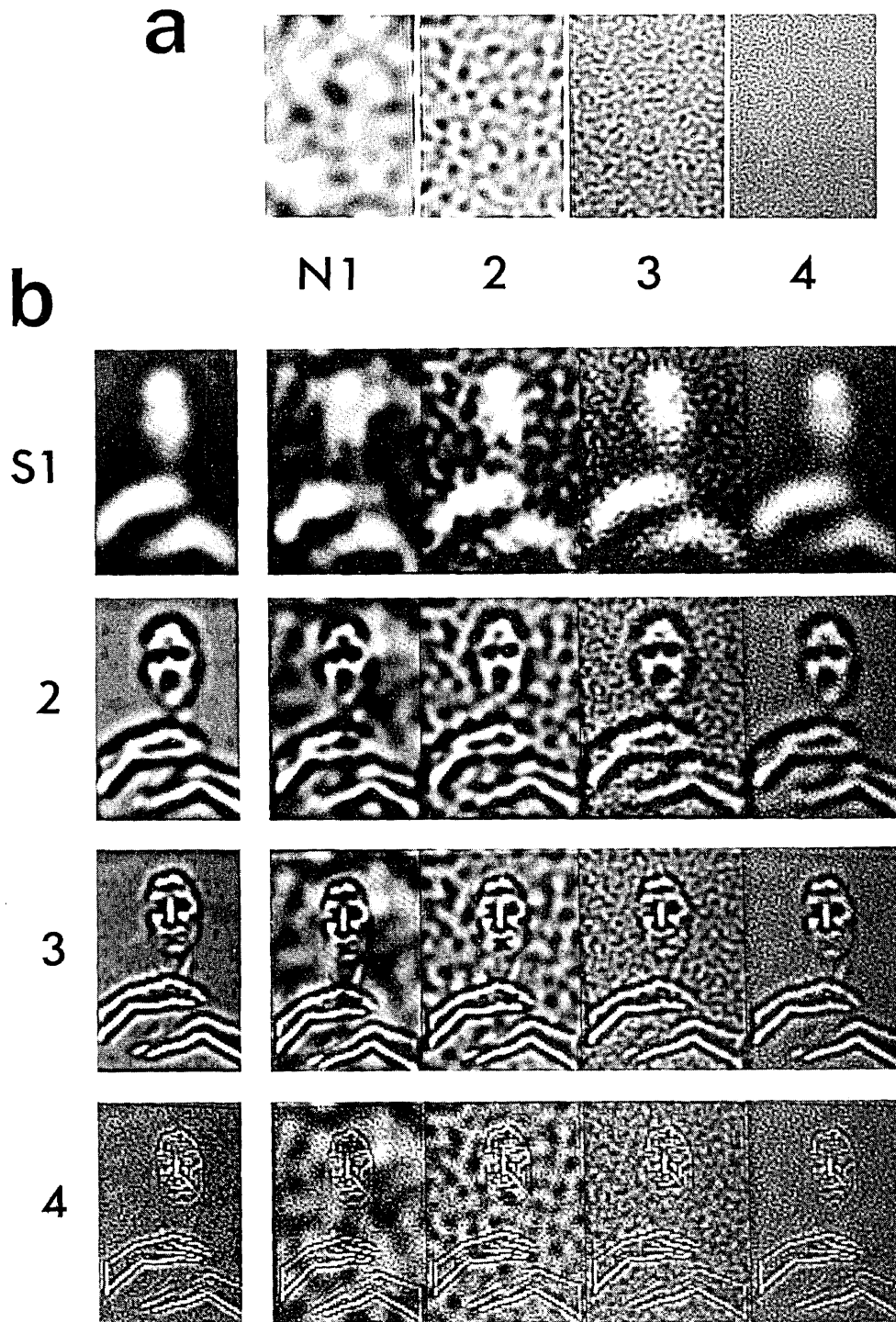


Fig. 5. Examples of all combinations of band-filtered signals plus band-filtered noise. a, Gaussian noise filtered in bands 1–4 (left to right). b, Band-filtered ASL signals plus band-filtered noise. Each row represents a single signal band with band 1 at the top and band 4 on the bottom. Each column (continuing downward from a) represents a single band of Gaussian noise. The leftmost column represents the noise-free signal.

estimation procedure. The single masking effectiveness parameter  $(s/n)_{50\%}$  used to describe each rating function is the  $s/n$  ratio at which the function attains 0.5 times its asymptotic height  $R_{\infty}$ .

Figure 7 shows the set of 16 estimated rating functions that describe the masking of each ASL band by each of the noise bands. The  $(s_i/n_j)_{50\%}$  values derived from the rating functions of Fig. 7 are graphically displayed in Fig. 8, which summarizes the cross-band-masking data. Bands 1, 2, and 4

mask themselves better than they mask any other band. Band 3 appears to mask band 4 slightly more than it masks itself, but we do not have a test of statistical significance for this effect.

*Masking as a Function of the Frequency Difference between the Test Stimulus and the Noise Masking Stimulus*  
Band 1 is more sensitive to masking by noise in its own band than are frequency bands 2, 3, and 4, which, when masking

themselves, are all equally effective; that is, let  $(s_i/n_j)_{50\%}$  represent the masking effectiveness of noise band  $j$  on signal band  $i$ . The points  $(s_i/n_j)_{50\%}$ ,  $i = 2, 3, 4$ , are all at the same level in Fig. 8; the points  $(s_1/n_1)_{50\%}$  is much higher.

To compare band 1 with the other bands, it is necessary to normalize the masking vulnerability of different bands. Masking vulnerability is indexed by self-masking  $(s_i/n_i)_{50\%}$ . The normalized masking effectiveness NME is

$$NME (s_i/n_j) = (s_i/n_j)_{50\%}/(s_i/n_i)_{50\%}$$

Masking as a function of the frequency separation between test and noise bands is illustrated in Fig. 9. The abscissa is the ratio  $f_s/f_n$  (on a log scale), where  $f$  represents the mean frequency of a band. The ordinate represents the log of the normalized masking effectiveness. The straight lines represent a mirror-symmetric function fitted to the

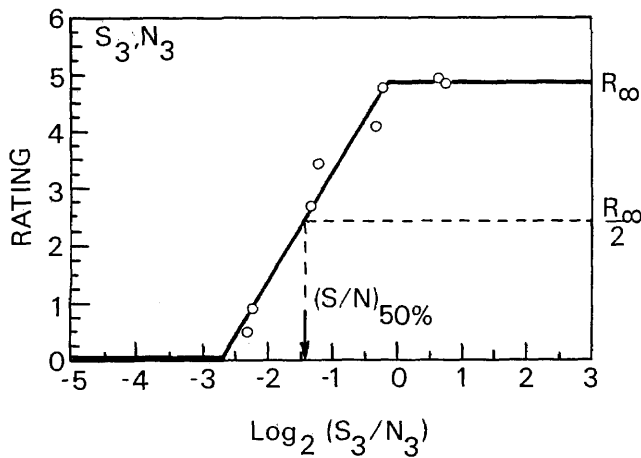


Fig. 6. Average ratings as a function of signal-to-noise ratio for the signal and the noise in band 3. The data are indicated by circles; the three-segment fit is indicated by the heavy lines. The dashed lines indicate the procedure for estimating  $(s/n)_{50\%}$ , the abscissa value under the arrow.

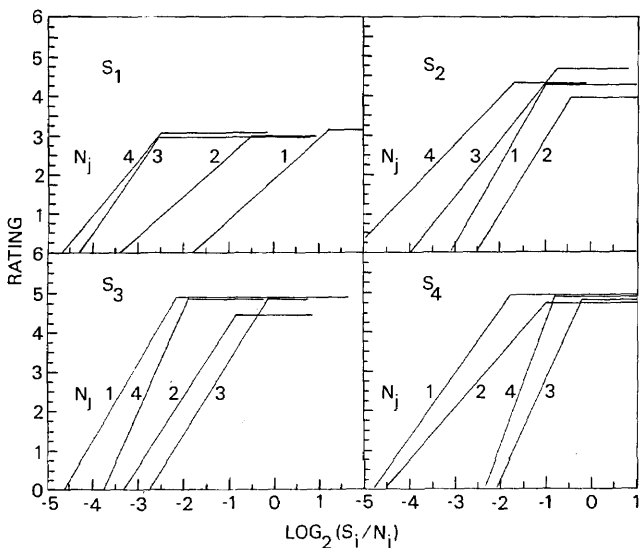


Fig. 7. Rating functions for cross-band masking. The abscissa is the signal-to-noise ratio; the ordinate is the mean rating, and the curves represent the three-segment best fits to the data. Each panel represents data from one signal band  $s_i$ ; the curve label indicates the band of the noise  $n_j$ .

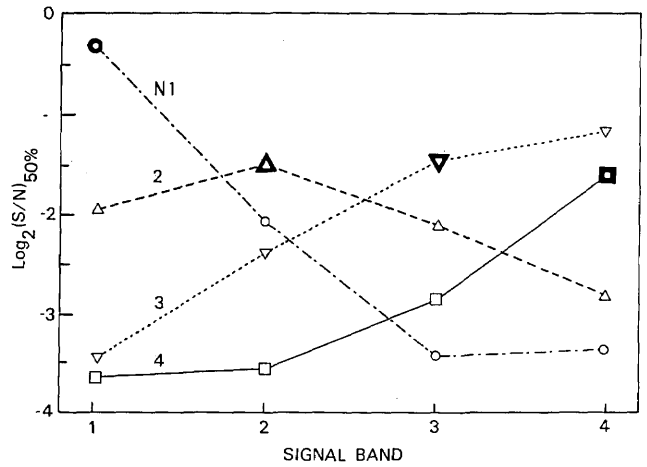


Fig. 8. Masking effectiveness of noise bands against signal bands. The abscissa is the signal band  $s_i$ ; the ordinate is the value of  $(s/n)_{50\%}$  derived from the rating functions (Fig. 7) by the estimation procedure shown in Fig. 6. The curve parameter indicates the noise band. Emphasized points indicate that the signal and the noise are in the same band.

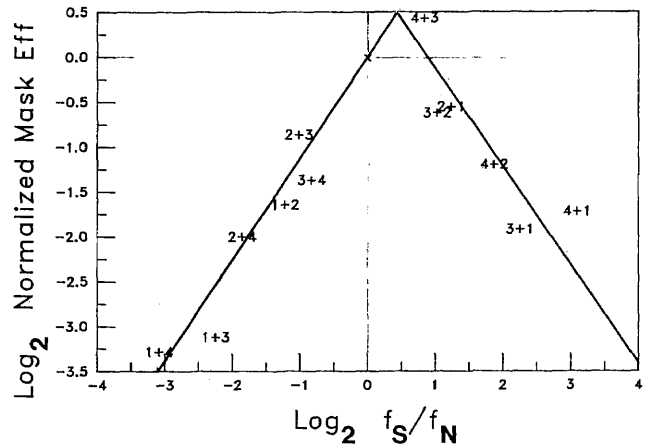


Fig. 9. Normalized cross-band masking as a function of frequency separation. Each band is represented by its mean frequency  $f$ . The abscissa represents the  $\log_2$  of  $f_{\text{signal}}/f_{\text{noise}}$ . The ordinate is the  $\log_2$  of the normalized masking effectiveness, the same data as in Fig. 8 with the curves for each signal band  $i$  moved up so that  $(s_i/n_i)_{50\%}$  falls at 0.0. Signal bands  $i$  and noise bands  $j$  are indicated by  $i + j$ ; the center of the + indicates the plotted datum. The straight lines represent the optimal mirror-symmetric fit to the data; the lines are centered above  $\log_2(f_s/f_n) = 0.46$  and with a slope of  $\pm 1.11$ .

data and constrained to pass through 0, 0. (The mirror-symmetric fit is the most convenient for determining whether there is any asymmetry between the masking effectivenesses of low and high frequencies.) The peak is located to the right of zero; the point of symmetry is  $x = \log_2(f_s/f_n) = 0.46$ , which represents a frequency ratio for optimal masking of 1:138. The slopes of the distance function are  $\pm 1.11$ .

Cross-band masking is quite adequately described in terms of log frequency separation ( $\log f_s - \log f_n$ ) without the necessity of referencing the particular frequencies that contribute to the separation. Masking falls off by a factor of slightly more than 2 when the frequency separation is doubled, a result that is generally consistent with data obtained with much simpler stimuli.<sup>12,13,15</sup> The right-of-center peak in Fig. 9 indicates that noise frequencies lower than the signal mask it slightly better than do frequencies higher than

the signal. This asymmetry is reflected in all six direct comparisons of masking of signal band  $i$  by noise band  $j$  compared with masking of signal band  $j$  and noise band  $i$ . For  $i > j$ , the masking effectiveness  $NME(s_i/n_j) > NME(s_j/n_i)$ . This masking asymmetry is opposite that obtained with data from simpler stimuli.<sup>20,21</sup>

Although masking falls off with increasing frequency distance between bands, with sufficient power, any noise band can obliterate any signal band; that is, in Fig. 7 all the rating functions were driven to zero at low signal-to-noise ratios. Our spatial-frequency filters are sufficiently narrow that this effect cannot be attributed to common-frequency masking, which occurs when frequencies in the tail of the noise happen to fall within the signal band and are so highly amplified that they change the signal-to-noise ratio within the signal band itself. Most masking between widely separated frequencies is caused by nonlinear distortion in the display system and the visual system, neither of which faithfully reproduces small-amplitude variations in large signals. Both systems, in effect, create masking noise at new frequencies when confronted with high-amplitude inputs. Indeed, the two extreme-left-hand and two extreme-right-hand points in Fig. 9 are at the intensity resolution limit of the display system and might have shown less masking effect (been lower in the figure) had the display system been better able to render small signal-to-noise ratios faithfully. To determine whether masking between widely separated frequencies also arises from genuine channel interactions would require bigger interactions than those observed here. All in all, the cross-band-masking data obtained with our complex displays are quite comparable with data obtained with sinusoidal gratings.

#### EXPERIMENT 4: ADDING SIGNALS FROM DIFFERENT BANDS

Typically, signal addition has been studied with simple, static signals at low contrast levels in which internal noise is dominant<sup>1,2,22,23</sup> rather than with realistic dynamic stimuli at high contrast levels with high levels of external noise. The purpose of experiment 4 is to discover quantitatively how ASL intelligibility is affected when two dynamic signals from different spatial-frequency bands are algebraically added. The effect on performance of adding two ASL signals is an inherently complex matter because it depends on the signal-to-noise level at which the addition is tested. This dependence is derived in part from the psychometric function (performance versus  $s/n$ ), which is concave up at low intensities and concave down at high intensities, and in part from more-complex factors. Thus, at high levels of  $s/n$ , performance cannot be improved by further increases in  $s$ . Insofar as we wish to characterize the efficiency of a detector in terms of internal noise, this would mean that at high input levels, internal noise is proportional to the input.<sup>24</sup>

At low levels of  $s$ , performance in detection tasks typically increases with the square of  $s$ ; i.e., power-law detection is obtained.<sup>25-27</sup> Square-law detection is consistent with constant internal noise, independent of  $s$ . Insofar as the square law also applies to band-limited ASL, doubling the amplitude of a signal in band  $i$  (and thereby quadrupling its power) might be expected to improve intelligibility more than would adding signal in band  $j$  (which would only double signal power).

In contrast, consider the addition of two signals at a high level of  $s/n$ . Within any single spatial-frequency band  $i$ , even with noiseless stimuli, performance is not so good as in the original unfiltered source images. Therefore, at a high signal level in band  $i$ , adding signals from another band  $j$  is more effective in improving performance than adding still more signal in band  $i$ . Thus different factors are critical for high-intensity and for low-intensity signal combinations, and their combinatorial effects are modeled by different rules.

To study how weak signals combine, we need a method of generating approximately equivalent weak signals. Weakening a signal by reducing the signal contrast relies on the observer's internal noise to weaken the signal. Adding external noise<sup>28</sup> is obviously the better way to control signal intelligibility. Pavel *et al.*<sup>24</sup> showed that for constant  $s/n$ , the signal contrast could be varied over a wide range without affecting intelligibility. Indeed, in a preliminary study (see Ref. 10, Exp. 4), this result was verified again with the current set of ASL stimuli. Thus, to study how signals combine, we may use any signals that fall within the enormous range of contrasts that is sufficient to overcome internal noise, and we vary intelligibility by varying external added noise.

#### Method

##### Overview

The first step in the procedure is to compose the spatial-frequency amplitude spectrum of an external noise stimulus so that it would mask all signal bands equally. Unfortunately, the rating functions in Fig. 7 are not parallel in the different signal bands, so equal masking of all spatial bands at different intensities is impossible with a single noise source. Given that limitation, we selected a particular noise stimulus to test, first, the intelligibility of weak signals in all bands  $i$  under this noise and, second, the intelligibility of all combinations of signals in band  $i$  with signals in band  $j$ .

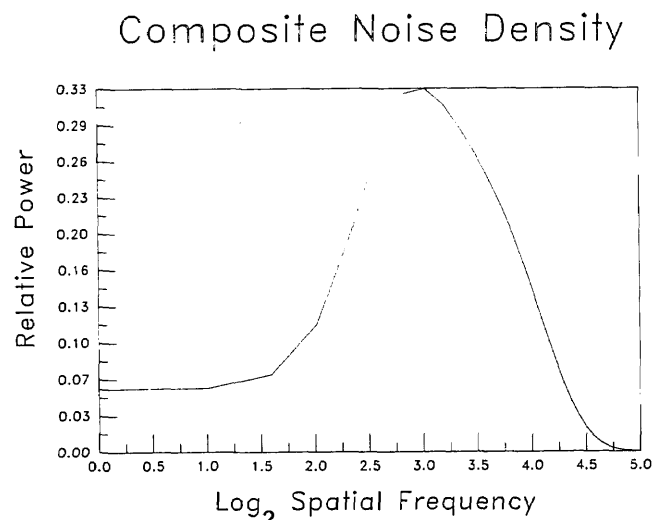


Fig. 10. Spatial power spectrum of the composite noise used in experiment 4. The abscissa is the  $\log_2$  of the spatial frequency in cycles per picture width ( $f_0$ , the width, is 64 pixels). The extreme-left-hand side represents 1 cycle per picture; the extreme-right-hand side represents 32 cycles per picture. The ordinate represents relative power on a linear scale.

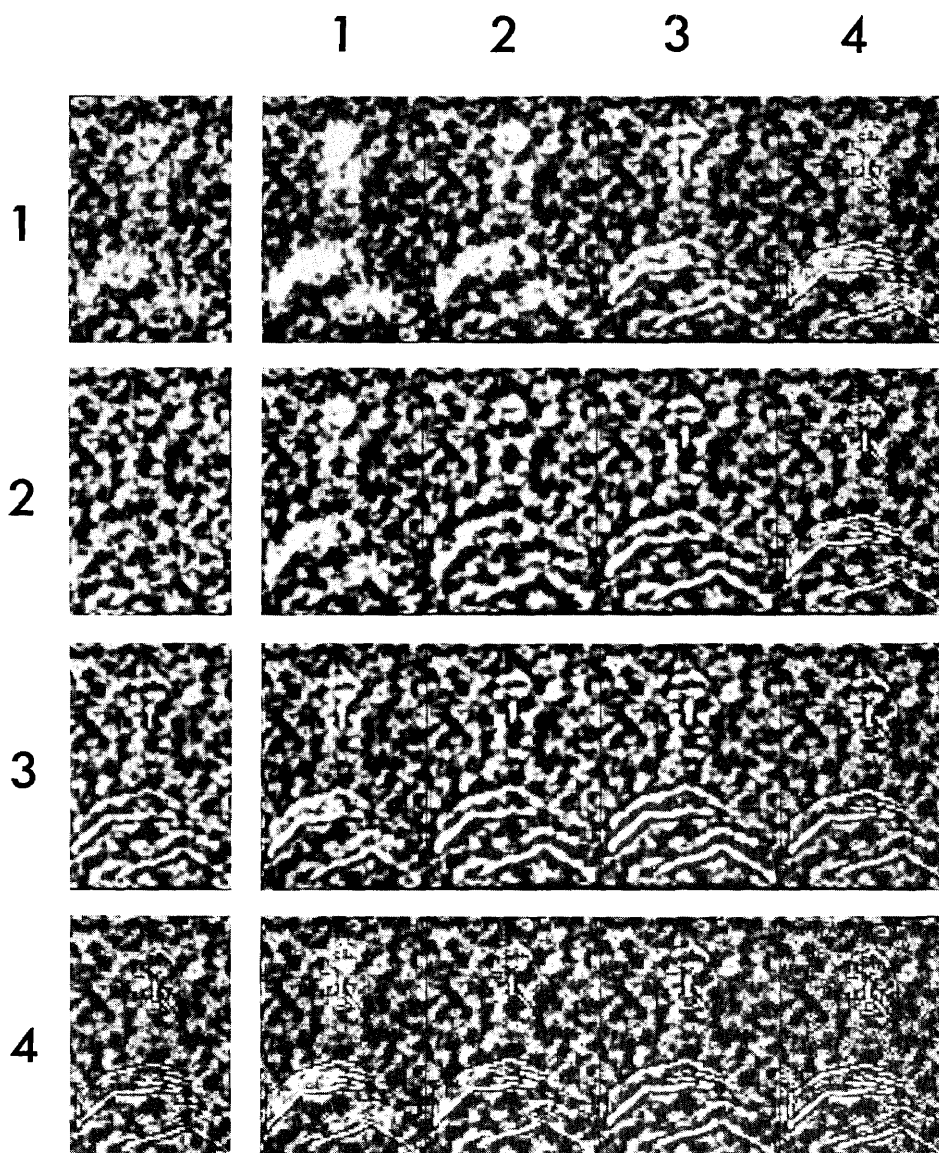


Fig. 11. Single frames illustrating the stimuli for experiment 4: The sum of weak signals in bands  $i$  and  $j$  plus the composite noise of Fig. 10. Composite noise is equally present in all stimuli. The leftmost column represents single-band signals, with the band indicated by the number at the left. The other panels represent stimuli composed of two signal bands, one component band indicated by the number at the left of the row and the other band indicated by the number at the top of the column.

#### Composite Noise

From the cross-band-masking data of experiment 3, we inferred a particular composite noise that would be expected to reduce weak signals in all bands to approximately equal intelligibilities. (We use the term composite noise to emphasize that the noise can be regarded as being composed of many spatial-frequency bands, each with a different amplitude and with a different temporal-frequency spectrum.) Full equality of intelligibility may be impossible with any composite noise because of the complex cross-band masking revealed in experiment 3. Figure 10 shows the spectrum of the noise that was used.

#### Signals

The signals were 80 ASL signs, basically the same set that was used in experiment 1b. They were produced at  $s_i/n = 0.25$ , where  $s_i$  indicates the amplitude of signal in band  $i$  and  $n$  indicates the rms amplitude of the composite noise stimu-

lus. All six combinations of signal in band  $i$  with signal band  $j$ ,  $j \neq i$ , were produced. There were four combinations of signal in band  $i$  with itself (i.e.,  $s/n = 0.5$ ) and four stimuli with signal in band  $i$  alone ( $s/n = 0.25$ ). Additionally, a composite signal was composed of the sum of all four bands represented by their amplitude in the  $s/n = 0.25$  condition. The composite signal was tested alone (the control condition) and under the composite noise (equivalent to  $s/n = 1.0$ ). The stimulus conditions are illustrated in Fig. 11.

#### Procedure

The 80 signs were divided to 16 blocks of 5 signs, balanced for difficulty. A Greco-Latin square design was used to generate a completely counterbalanced design in which every block of ASL signs occurred in every signal condition, and the order of conditions was balanced over subjects. This required generating 16 different hour-long stimulus tapes, one for each of the 16 subjects run in this experiment.



The viewing and testing conditions were similar to those described for experiment 1 and particularly for experiment 1b. Subjects were fluent ASL signers from the community. As before, all subjects had good vision under the experimental conditions as determined by an acuity test administered before the experiment.

**Results**

Figure 12 shows the results for all classes of signals confined to a single band. At  $s/n = 0.25$ , intelligibility in all bands is below 9%. At  $s/n = 0.5$ , intelligibility in bands 1 and 2 is 17.5%, whereas performance in bands 3 and 4 is 60.0%. At  $s/n = \infty$ , the conditions run to test the filters in experiment 1b, intelligibility rises to 66.4% in the lowest band and up to 87.5% in band 3.

Figure 13 shows the same data as Fig. 12 plus the six additional summation conditions of band  $i$  with band  $j, i \neq j$ . The points indicated with circles in Fig. 13 are precisely the same  $s/n = 0.5$  points as in Fig. 12. Since they do not seem to fall any differently on the curves than do nearby points that represent different bands, it appears that summation is quite similar within and between bands.

*Statistical Analysis of the Data*

The design of experiment 4 involves three factors: 16 conditions  $\times$  16 subjects  $\times$  16 stimulus sets. Because each subject saw each stimulus set only once (and not once in each condition), only 256 of the 4096 possible conditions were run. Typical analysis-of-variance designs are inappropriate for

such a sparse design, so a simple linear model was developed. A subject's score  $y$  for a set of five stimulus items that constitute a condition ranges from 0 to 5 and is assumed to be the sum of five terms: the grand mean  $m$ , factors for condition difficulty  $c_i$ , the subject's skill  $s_j$ , the ASL set difficulty  $a_k$ , and finally a term representing random error  $\epsilon_{i,j,k}$ :

$$y_{i,j,k} = m + c_i + s_j + a_k + \epsilon_{i,j,k}$$

Condition difficulty  $c_i$  is estimated by

$$c_i = \frac{1}{16} \sum_{j=1}^{16} y_{i,j,k} - m = \frac{1}{16} \sum_{k=1}^{16} y_{i,j,k} - m,$$

that is, by averaging over all subjects and stimulus sets in which condition  $i$  occurred and subtracting  $m$ . Factors  $s_j$  and  $a_k$  are estimated similarly. The variance  $\sigma^2$  of the random error  $\epsilon$  is  $(1/210) \sum \epsilon^2$ , where 210 represents the degrees of freedom, the number of cells (256) reduced by the number of estimated parameters (1 + 15 + 15 + 15).

The rms error  $\sigma$  was found to be 0.984. This is approximately what would be predicted from the binomial variability of the data if the predictions  $\hat{y}_{i,j,k} = c_i + s_j + a_k$  were based on a completely correct model. The standard error of the mean of the scores shown in Figs. 12 and 13 is  $\pm 4.92\%$ .

*Summation as a Function of Frequency Distance between Bands*

The amount of intelligibility summation as a function of the frequency separation between component signals can be

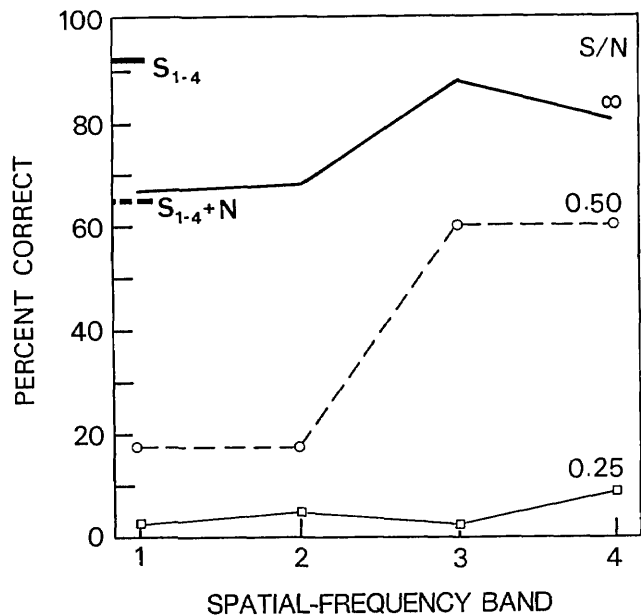


Fig. 12. Data from experiment 4: Intelligibility of band-limited single-band signals in composite noise. The abscissa indicates the band of the signal; the ordinate indicates the percent correct scored by the 16 subjects in the intelligibility test. The curve parameter indicates the signal-to-noise ratio of the stimuli. The curve labeled  $\infty$  represents data obtained without added noise in experiment 1 (with different subjects and a slightly different stimulus set). On the left-hand ordinate, the point  $S_{1-4}$  indicates intelligibility of the noise-free sum signal of band 1 + band 2 + band 3 + band 4; the point  $S_{1-4}+N$  indicates the intelligibility of the same signal plus noise ( $s/n = 1$ ).

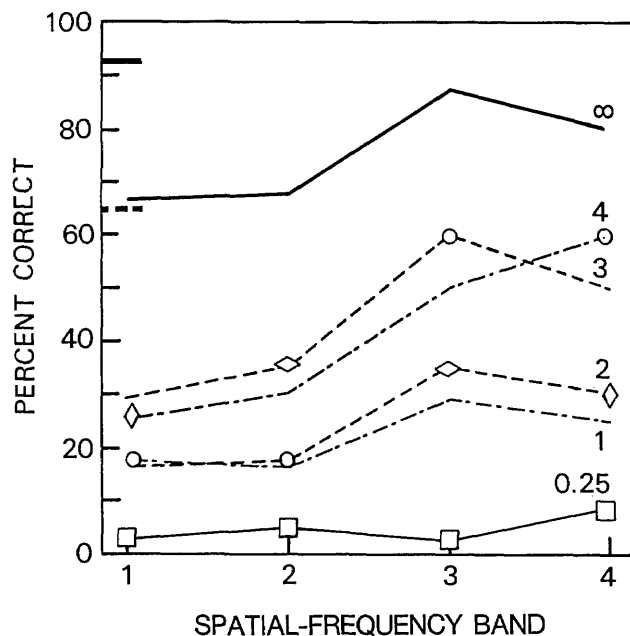


Fig. 13. Data from experiment 4: intelligibility of pairs of band-limited signals in composite noise. The ordinate, the abscissa, and the curves labeled 0.25 and  $\infty$  are as in Fig. 12. The dashed curves indicate signals composed of band  $i$  (indicated on abscissa) and band  $j$  (indicated as the curve parameter). The open circles represent data for  $i = j$ , the middle curve of Fig. 12. The flat diamonds represent the addition of nearby signal bands (2 and 3); the tall diamonds represent the addition of distant bands (1 and 4). The pairs indicated by diamonds are matched for the strengths of their constituent signals.

tested nicely by using the data of experiment 4. Because, at  $s/n = 0.5$ , bands 1 and 2 have, by coincidence, exactly the same intelligibility (17.5%) and bands 3 and 4 have the same intelligibility (60.0%), we compare the intelligibility of band 1 plus band 4 (wide separation) with that of band 2 plus band 3 (small separation). These two points are at slightly different intelligibility levels in Fig. 13; the small band separation (flat diamonds) at 35% is somewhat higher than the large separation (thin diamonds) at 25%. The probability that a difference this large would occur by chance, estimated by a one-tail  $z$  test, is 0.040.

To determine whether it is more efficient to improve a weak signal in band  $i$  by adding more energy in  $i$  or do so by adding energy in an adjacent band  $j$ , we compare the effects of summing two signals at  $s/n = 0.25$ . In Fig. 13, the crossings of the curves labeled 3 and 4 at the extreme right and the crossings of the curves labeled 1 and 2 at the extreme left indicate that there is a tendency for the sum of band 4 + band 4 ( $I = 60\%$ ) and of band 3 + band 3 ( $I = 60\%$ ) to be more intelligible than band 3 + band 4 ( $I = 50\%$ ) and for the sums band 1 + band 1 and band 2 + band 2 (both  $I = 17.5\%$ ) to be slightly more intelligible than band 1 + band 2 ( $I = 16.3\%$ ). The probabilities of these differences' occurring under the null hypothesis are 0.024 and 0.209, respectively. Taken together, these observations imply that, with the signal levels studied here, there is a small but occasionally significant tendency for component signals to contribute more to intelligibility when they are closer in frequency.

#### Efficiency When Signal Power Is Constrained

For practical purposes, when two different weak visual ASL signals are summed, the effect of frequency separation on intelligibility is small. All the factors that might have contributed to a separation effect or an inverse separation effect are almost in balance at the  $s/n$  values investigated here. To improve intelligibility, given a signal in band  $i$ , adding more signal in any other band  $j$  is almost as effective as adding more signal in  $i$ . In these signal manipulations, we are speaking of signal amplitudes. If we were concerned with signal power rather than with rms amplitude, then it would clearly be more efficient to distribute the power over different bands. Doubling the amplitude within a band quadruples the power, whereas the power of signals in disjoint bands adds linearly.

## SUMMARY AND CONCLUSIONS

(1) In low-resolution dynamic ASL images ( $96 \times 64$  pixels), it is possible to divide the original signal into four different frequency bands, each of which is quite intelligible (67–87% for isolated ASL signs) and each of which could serve for ordinary ASL communication.

(2) The empirically determined temporal-frequency spectrum of ASL is approximately the same in all spatial-frequency bands.

(3) The ratio of root-mean-square signal amplitude to noise amplitude,  $s/n$ , at which ASL becomes intelligible is nearly the same for the three highest bands, but the critical  $s/n$  is higher for the lowest-frequency band.

(4) The masking of signals in one band by noise in another is governed simply by the ratio of frequencies between the bands (the difference of the log frequencies). There is

asymmetry: noise lower in spatial frequency than the signal is more effective in masking than is higher-frequency spatial noise. When the frequency separation between signal and noise is increased by a factor of 2, intelligibility can be maintained at 1/2 the original signal-to-noise ratio.

(5) When two weak signals ( $s/n = 0.25$ ) are added, the intelligibility of the summed signal is slightly greater when the two signals are in adjacent frequency bands than when they are widely separated bands; and intelligibility is slightly greater when the two signals are identical than when they are in adjacent bands. If the signal power—not amplitude—is limited, intelligibility is maximized by dispersing the signal power widely across frequency bands.

## APPENDIX A: FILTER-GENERATION ALGORITHM

This algorithm generates  $K$  filters that divide frequency space ( $\omega_x$  and  $\omega_y$ ) into partially overlapping annular regions whose boundaries are adjustable. The summed output of all the filters equals the original input signal.

Let  $K$  be the desired number of filters. Let LP represent the Fourier transform of a low-pass filter; that is,  $|| LP(\omega_x, \omega_y) ||$  is monotonically decreasing in  $\omega_x$  and  $\omega_y$ . (The particular LP $_i$  that are used to feed the algorithm are defined below.) We use the terms center and surround analogously to their use in composing difference-of-Gaussian filters; they refer to  $x, y$  spread functions of the filters. The center and surround components are used as kernels to generate the filters. The surround of filter  $K - i + 1$  becomes the center of filter  $K - i$  (the next lower filter in terms of frequency). In the sum of all the filters, all the centers and surrounds cancel, and the original source image is recovered. The steps in the algorithm are stated in terms of the two-dimensional Fourier transforms of the filters and their components:

(1) Define  $F_K$ , the highest-frequency filter. The center of  $F_K$  is defined to be  $C_K = 1$ . The surround of  $F_K$  is defined in terms of LP $_K$  (see below) as  $S_K = 1 - (1 - LP_K)^m$ ; then the  $K$ th filter is  $F_K = C_K - S_K = (1 - LP_K)^m$ .

(2) Do the following loop  $K - 2$  times ( $i = 1, K - 2$ ) to generate, in sequence, the filters  $K - 1, K - 2, \dots, 2$ :

(a) Define the center of the  $K - i$  filter as the surround of the previously defined filter:  $C_{K-i} = S_{K+i-i}$ .

(b) Define the surround of the  $K - i$  filter:  $S_{K-i} = 1 - (1 - LP_{K-i})^m$ . The surround is a low-pass filter derived from a generating low-pass filter LP $_{K-i}$  chosen so that  $S_{K-i}$  will have a lower cutoff frequency than  $C_i$  in accordance with the desired partition of frequency space.

(c) Define the  $K - i$  filter as the center minus the surround:  $F_{K-i} = C_{K-i} - S_{K-i}$ .

(d) Increase  $i$ ; if  $i \leq K - 2$ , return to step (a); otherwise, continue to step (3).

(3)  $F_1 = 1 - \sum_2^K F_i = S_2$ ; that is,  $F_1$  is the low-pass filter that was chosen as the surround of  $F_2$ ; it encompasses all the residual signal. Note that  $\sum_1^K F_i = 1$ .

To begin the algorithm with  $F_K = (1 - LP_K)^m$ , LP $_K$  must be defined. Let LP $_K$  be a two-dimensional Gaussian low-pass filter whose frequency-domain representation is

$$LP_K(\omega_x, \omega_y, \sigma_x, \sigma_y) = \exp[-2\pi^2(\sigma_x^2\omega_x^2 + \sigma_y^2\omega_y^2)],$$

where  $\omega_x$  and  $\omega_y$  are the frequency components in the  $x$  and  $y$  directions, respectively, and  $\sigma_x$  and  $\sigma_y$  are the  $x$  and  $y$  widths of the generating spatial Gaussians. Since  $F_K = (1 - F_K)^m$ , as  $m$  increases, the frequency cutoffs become steeper, and the overlap between filters is reduced (which is good); but for  $m > 4$ , the ringing in the  $x$ - $y$ -space domain becomes obtrusive (which is bad). Therefore  $m = 4$  was chosen.

## ACKNOWLEDGMENTS

This research was supported partly by National Science Foundation, Science and Technology to Aid the Handicapped, grant no. PFR-80171189; and by U.S. Air Force Life Sciences Directorate grant AFOSR-85-0364. The work was conducted at New York University and partially fulfilled the requirements for a Ph.D. degree in experimental psychology for Thomas R. Riedl. Throughout the work, we received valuable advice and guidance from Misha Pavel and Michael S. Landy. We thank James P. Thomas for his comments on the manuscript and Geoffrey Iverson for statistical advice. We express our gratitude for the help that we have received from many individuals in the deaf community and from organizations, including The Disabled Student Services Office of New York University, The New York Society for the Deaf, and D.E.A.F. We thank our signer, Ellen Roth, and we wish to acknowledge the skillful technical assistance of Robert Picardi and August Vanderbeek.

\* Present address, AT&T Bell Laboratories, Whippany Hill, New Jersey 07981; requests for reprints should be addressed here.

## REFERENCES

1. For a succinct review, see N. Graham, "Detection and identification of near-threshold visual patterns," *J. Opt. Soc. Am. A* **2**, 1468-1482 (1985).
2. For a review, see L. A. Olzak and J. P. Thomas, "Seeing spatial patterns," in *Handbook of Perception and Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas, eds. (Wiley, New York, 1986), Chap. 7.
3. G. E. Legge, D. G. Pelli, G. S. Rubin, M. M. Schleske, "Psychophysics of reading—I. Normal vision," *Vision Res.* **25**, 239-252 (1985).
4. D. H. Parish and G. Sperling, "Object spatial frequency, not retinal spatial frequency, determines identification efficiency," *Invest. Ophthalmol. Vis. Sci. Suppl.* **28**, 359 (1987).
5. G. Sperling and D. H. Parish, "Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of discrimination," in *Mathematical Studies in Perception and Cognition* (Department of Psychology, New York University, New York, N.Y., 1987).
6. J. D. Schein and M. T. Delk, Jr., *The Deaf Population of the United States* (National Association of the Deaf, Silver Spring, Md., 1974).
7. G. Sperling, "Bandwidth requirements for video transmission of American Sign Language and finger spelling," *Science* **210**, 797-799 (1980).
8. G. Sperling, "Video transmission of American Sign Language and finger spelling: present and projected bandwidth requirements," *IEEE Trans. Commun.* **COM-29**, 1993-2002 (1981).
9. G. Sperling, M. Landy, Y. Cohen, and M. Pavel, "Intelligible encoding of ASL image sequences at extremely low information rates," *Comput. Vision Graph. Image Process.* **31**, 335-391 (1985).
10. T. R. Riedl, "Spatial frequency selectivity and higher level human information processing," doctoral dissertation (New York University, New York, N.Y., 1985).
11. P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.* **COM-31**, 532-540 (1983).
12. G. B. Henning, B. G. Hertz, and J. L. Hinton, "Effects of different hypothetical detection mechanisms on the shape of spatial-frequency filters inferred from making experiments. I. Noise masks," *J. Opt. Soc. Am.* **71**, 574-581 (1981).
13. G. Legge and J. Foley, "Contrast masking in human vision," *J. Opt. Soc. Am.* **70**, 1458-1471 (1980).
14. S. Stecher, C. Sigel, and R. V. Lange, "Composite adaptation and spatial frequency interactions," *Vision Res.* **13**, 2527-2531 (1973).
15. C. F. Strohmeyer III and B. Julesz, "Spatial-frequency masking in vision: critical bands and spread of masking," *J. Opt. Soc. Am.* **62**, 1221-1232 (1972).
16. C. F. Strohmeyer III, S. Klein, B. M. Dawson, and L. Spillmann, "Low spatial-frequency channels in human vision: adaptation and masking," *Vision Res.* **22**, 225-233 (1982).
17. D. J. Tolhurst, "Adaptation to square-wave gratings: inhibition between spatial frequency channels in the human visual system," *J. Physiol.* **226**, 231-248 (1972).
18. H. R. Wilson and J. R. Bergen, "A four mechanism model for threshold spatial vision," *Vision Res.* **19**, 19-32 (1979).
19. J. P. Chandler, "STEPT," in *Quantum Chemistry Program Exchange* (Department of Chemistry, Indiana University, Bloomington, Ind., 1965).
20. J. Nachmias and A. Wever, "Discrimination of simple and complex gratings," *Vision Res.* **15**, 217-223 (1975).
21. D. J. Tolhurst and L. P. Barfield, "Interactions between spatial frequency channels," *Vision Res.* **18**, 951-958 (1978).
22. N. Graham and J. Nachmias, "Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple models," *Vision Res.* **11**, 251-259 (1971).
23. N. Graham, "Psychophysics of spatial-frequency channels," in *Perceptual Organization*, M. Kubovy and J. Pomerantz, eds. (Erlbaum Halstead, Potomac, Md., 1980), pp. 1-25.
24. M. Pavel, G. Sperling, T. Riedl, and A. Vanderbeek, "Limits of visual communication: the effects of signal-to-noise ratio on the intelligibility of American Sign Language," *J. Opt. Soc. Am. A* **4**, 2355-2365 (1987).
25. C. R. Carlson and R. W. Klopfenstein, "Spatial-frequency model for hyperacuity," *J. Opt. Soc. Am. A* **2**, 1747-1751 (1985).
26. J. Nachmias and R. V. Sansbury, "Grating contrast: discrimination may be better than detection," *Vision Res.* **14**, 1039-1042 (1974).
27. C. F. Strohmeyer and S. Klein, "Evidence against narrow-band spatial frequency channels in human vision: the detectability of frequency modulated gratings," *Vision Res.* **15**, 899-910 (1975).
28. D. G. Pelli, "Effects of visual noise," doctoral dissertation (University of Cambridge, Cambridge, 1981).