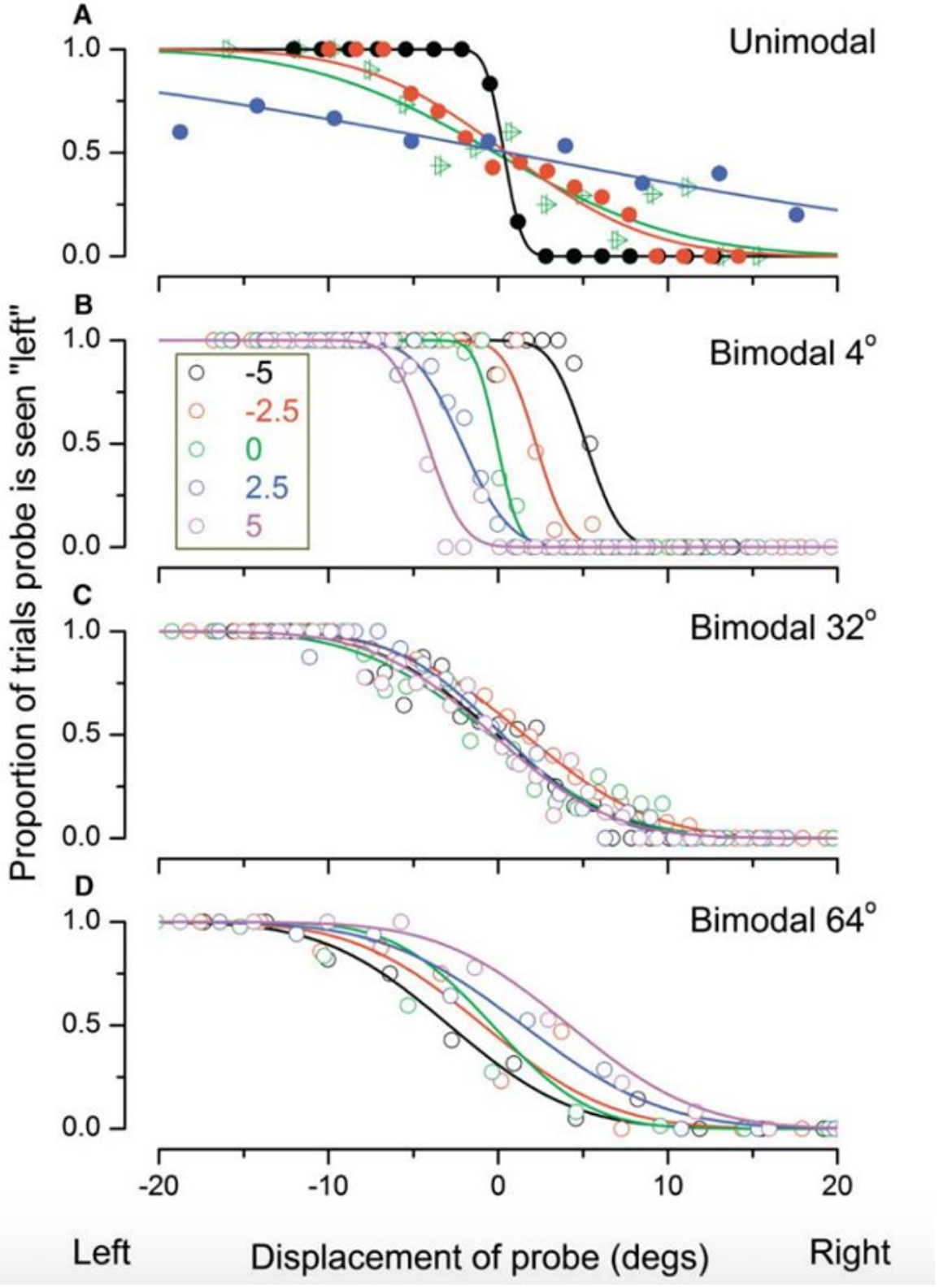
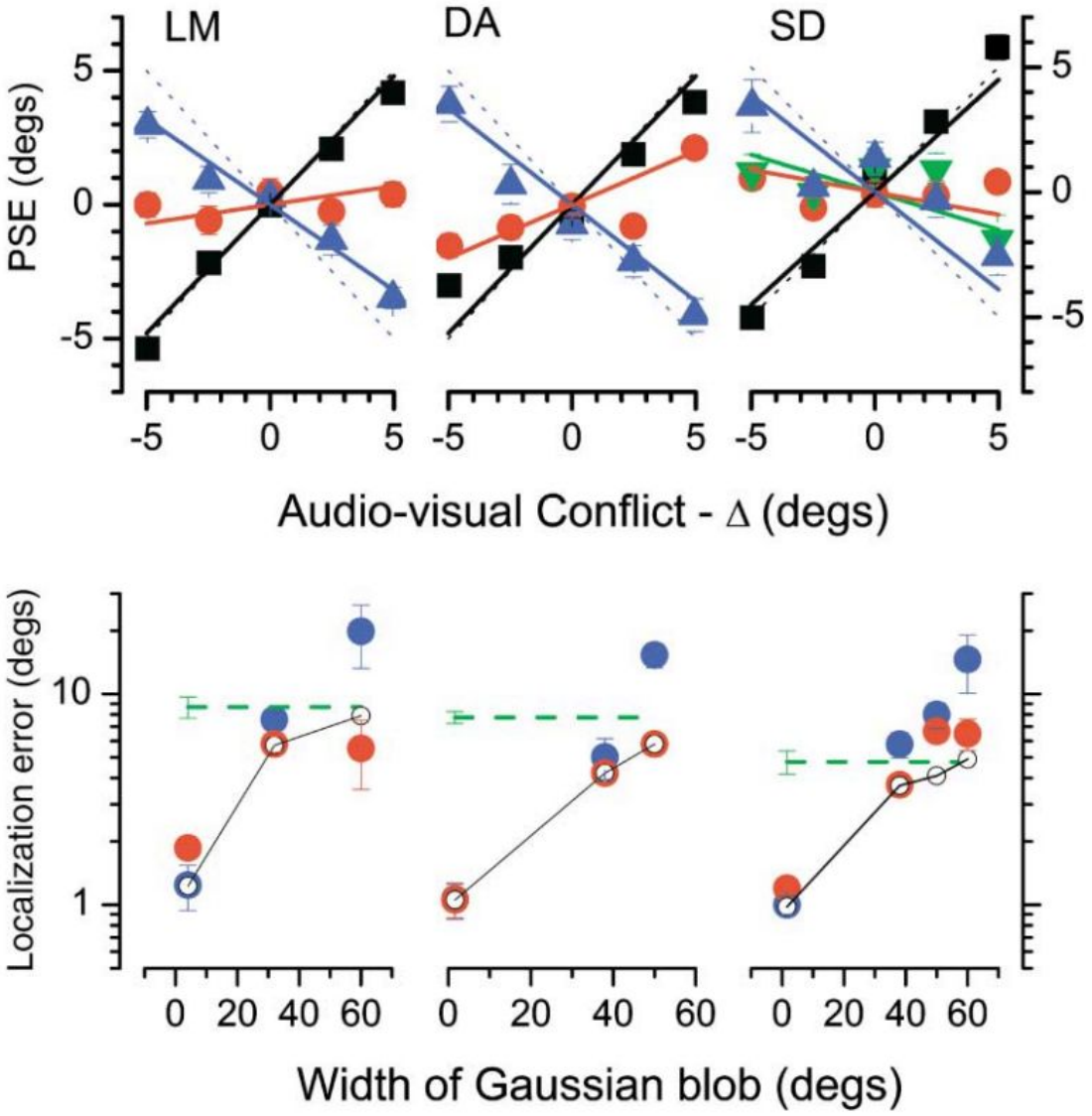
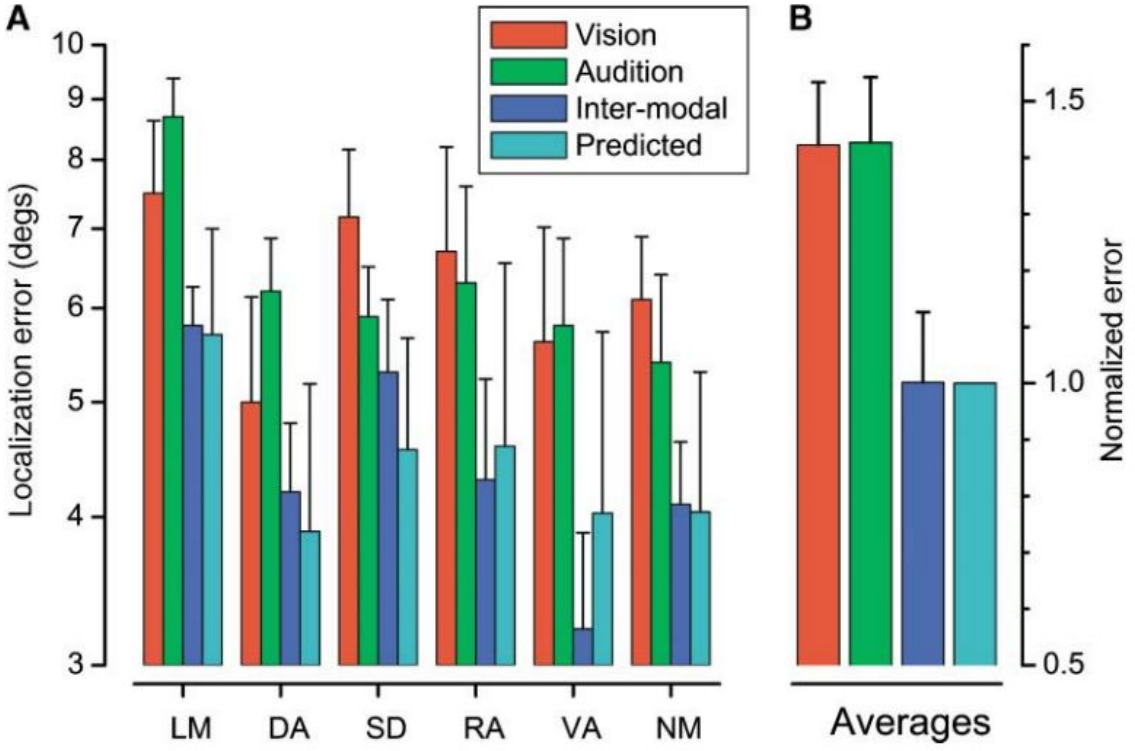

A
Unimodal
B
Bimodal 4°
C
Bimodal 32°
D
Bimodal 64°
1.0
0.5
0.0
-5
-2.5
0
2.5
5
Proportion of trials probe is seen "left"
-20
-10
0
10
20
Left
Displacement of probe (degs)
Right


LM
DA
SD
PSE (degs)
5
0
-5
-5
0
5
Audio-visual Conflict - Δ (degs)
Localization error (degs)
10
1
0
20
40
60
Width of Gaussian blob (degs)

$$\hat{S}_{VA} = w_V \hat{S}_V + w_A \hat{S}_A \qquad w_A = \frac{1/\sigma_A^2}{1/\sigma_A^2 + 1/\sigma_V^2} = \frac{\sigma_V^2}{\sigma_A^2 + \sigma_V^2}$$



The Ventriloquist Effect Results from Near-Optimal Bimodal Integration

Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. Current Biology, 14(3), 257-262. doi:10.1016/j.cub.2004.01.029

Presented by: Le Nhat Hung (11:30-13:00), Anuradha Mallik (13:00-14:30)

Figure 1. Unimodal and Bimodal Localization of Visual and Auditory Stimuli

(A) Psychometric functions for localizing either an auditory click (green speaker-shaped symbols) or visual blobs of various Gaussian space constants ($2\sigma = 4°$, black; $2\sigma = 32°$, red; or $2\sigma = 64°$, blue). The auditory stimuli were brief (1.5 ms) clicks, with their apparent position on the median plane controlled by interaural time differences (ITDs). Timing resolution was 15.3 μs, corresponding to about 1.2° in lateral angle (calibrated separately for each observer). All trials comprised two stimulus presentations, one presented near-centrally (with a small jitter from trial to trial) and the other displaced leftward or rightward by an amount given by the abscissa. The ordinate shows the proportion of times the observer judged the probe presentation (randomly first or second) "leftward." Two hundred fifty trials were run for each condition, over ten separate sessions. The data show that the points of subjective alignment are similar for all stimuli (≈0°), while the widths of the visual functions (assumed to reflect internal neural noise) increase with the width of the visual stimulus (see also Figure 2B). The width of the auditory function lies midway between the smallest and largest visual stimuli.

(B–D) Psychometric functions for localizing bimodal presentations of the click and blob together (click centered within the blob), for blob widths 4° (B), 32° (C), or 64° (D). One presentation (randomly first or second) was the conflict stimulus, with the visual stimulus horizontally displaced Δ° rightward and the auditory stimulus displaced Δ° leftward. In the other non-conflict presentation, both stimuli were displaced in the same direction by the amount shown by the abscissa (positive indicates rightward). The values of Δ were −5° (black symbols), −2.5° (red), 0° (green), +2.5° (blue), or +5° (mauve), randomly intermingled within each session. Two hundred fifty trials were run for each condition, over ten separate sessions. Figure 2 summarizes these data, together with those of another two subjects.

Figure 2. Effect of Conflict on Localization of Visual and Auditory Stimuli

(A) PSE for localizing the bimodal stimulus as a function of audio-visual conflict (Δ). The PSE was defined as the median of the cumulative Gaussian distributions for each observer (see examples of Figure 1). The results are shown for various sizes of visual blob, indicated by the colors (L.M.: 4°, black squares; 32°, red circles; and 64°, blue triangles; D.A.: 1.6°, 38°, and 50°; S.D.: 1.6°, 38°, and 50°, green triangles; and 64°). For the small blob sizes, the perceived bimodal position varied directly with the visual stimulus, while for the larger blobs they varied with audition. For the mid-sized blobs, there is little variation. The dotted black and blue lines show the predicted result if vision and audition were to dominate totally. The continuous lines are not fits to the data, but predictions from optimal statistical combination of cues (Equation 1), obtained solely from the parameters of the unimodal measurements (shown in [B]).

(B) Localization error (given by the root-variance of the psychometric functions) as a function of blob size. The blue symbols show the unimodal visual thresholds and the green dashed line the auditory unimodal thresholds. The bimodally measured thresholds are shown by the filled red circles, and the predicted thresholds by the black open circles joined by the lines. Note that for all subjects, the second point along the abscissa is where visual and auditory localization errors are most similar. In these cases, the model predicts that bimodal variance should be lower than those of either vision or audition (see also Figure 3).

Figure 3. Comparison of Actual and Predicted Thresholds for Visuo-Auditory Localization

(A) Visual, auditory, bimodal, and predicted localization errors for six subjects, for the visual condition which yielded thresholds most similar to the auditory thresholds. Visual blobs of 38° were used in all cases except for L.M., in which they were 32°. This graph includes data from the three subjects of Figure 2B. The bars show standard errors of the mean, calculated by 500 repetitions of a bootstrap procedure, with the error of the predicted result calculated from the two unimodal thresholds. The probabilities of rejecting the null hypothesis (that the bimodal threshold was the same as the best unimodal threshold), calculated by the bootstrap t test (one-tailed, 5000 repetitions) for the six observers were: L.M., 0.24; D.A., 0.5; S.D., 0.11; R.A., 0.057; V.A., 0.03; and N.M., 0.07.

(B) Averages of all six subjects, after normalizing to unity for the predicted value (so the error in that condition is zero). The improvement of the bimodal condition over the average unimodal conditions was 1.425, compared with a predicted value of 1.414. Averaging the performance of the six observers and comparing best unimodal performance with bimodal performance in a one-tailed, matched samples t test produced a clearly significant difference, with $t_5 = 3.69$ and $p < 0.01$. When the bimodal thresholds were tested against the predicted thresholds by the bootstrap t test technique, none were significantly different (all p values greater than 0.2).

Equation 1. Optimum combination of the independent auditory and visual estimates $\widehat{S}_A$ and $\widehat{S}_A$

Several authors have suggested that multimodal information may be combined in an optimal way by summing the independent stimulus estimates from each modality according to an appropriate weighting scheme. The weights correspond to the inverse of the noise associated with each estimate, given by the variance σ^2 of the underlying noise distribution (assumed to be approximated by the squared width of the psychometric function). This model is “optimal” in that it combines the unimodal information to produce a multimodal stimulus estimate with the lowest possible variance (that is, with the greatest reliability).

Equation 2. Relative weights for each modality w_A and w_V

and likewise for w_V. w_A and w_V are the relative weights for each modality, inversely proportional to their localization variances. Estimates of σ_V^2 and σ_A^2 can be obtained by the Gaussian fit of the unimodal data of Figure 1A (also plotted in Figure 2B).