Factors influencing audiovisual fission and fusion illusions

Tobias S. Andersen*, Kaisa Tiippana, & Mikko Sams

Laboratory of Computational Engineering, Helsinki University of Technology, Finland

P.O. Box 3000, 02015 HUT, Finland

Accepted for publication in Cognitive Brain Research 2004

*corresponding author: tobias@lce.hut.fi

Abstract

Information processing in auditory and visual modalities interacts in many circumstances. Spatially and temporally coincident acoustic and visual information are often bound together to form multisensory percepts [13,16]. Shams and coworkers recently reported a multisensory fission illusion where a single flash is perceived as two flashes when two rapid tone beeps are presented concurrently [11,12]. The absence of a fusion illusion, where two flashes would fuse to one when accompanied by one beep, indicated a perceptual rather than cognitive nature of the illusion. Here we report both fusion and fission illusions using stimuli very similar to those used by Shams et al. By instructing subjects to count beeps rather than flashes and decreasing the sound intensity to near threshold we also created a corresponding visually induced auditory illusion.

We discuss our results in light of four hypotheses of multisensory integration, each advocating a condition for modality dominance. According to the *discontinuity hypothesis* [12], the modality in which stimulation is discontinuous dominates. The *modality appropriateness hypothesis* [16] states that the modality more appropriate for the task at hand dominates. The *information reliability hypothesis* [10] claims that the modality providing more reliable information dominates. In strong forms, none of these three hypotheses applies to our data. We re-state the hypotheses in weak forms so that discontinuity, modality appropriateness and information reliability are factors which increase a modality's tendency to dominate. All these factors are important in explaining our data. Finally, we interpret the effect of instructions in light of the *directed attention hypothesis* which states that the attended modality is dominant [16].

Theme: Sensory Systems

Topic: Multisensory

Keywords: multisensory illusions; discontinuity hypothesis;
modality appropriateness; information reliability; directed attention; illusory flashes

1. Introduction

One of the challenges of cognitive neuroscience is to reveal how information from the multiple sensory modalities is integrated. Evolution has designed multisensory integration to take advantage of situations where different senses provide information on the same object. As an example, speech perception is facilitated by viewing a congruent articulating face [2,3,7,14]. Also, detection of weak auditory and visual stimuli improves drastically when they occur in unison [13]. In contrast, illusory percepts often occur when incongruent information is presented to two modalities. This is seen in the McGurk effect [4,6,8,9] where viewing an incongruently articulating face categorically alters the auditory speech percept. Multisensory illusions have two important characteristics. First, their effect is often greater than that of congruent stimuli. In the McGurk effect, watching an incongruent utterance can change the category of *all* of a subjects' responses. In comparison, the improvement of identification obtained from watching congruent speech is smaller. Therefore the incongruent case often affords a stronger test for hypotheses on multisensory integration. Second, in multisensory integration, one modality often dominates the other. This effect is likely to occur also when stimulation is congruent, but only in the incongruent case can we relate subjects' perception to one modality more than the other. Hence, multisensory illusions are far from mere curiosities, but provide an excellent opportunity to test hypotheses on the mechanism of integration.

Recently, Shams and coworkers reported a new multisensory illusion where a single flash is perceived as two when two rapid tone beeps are presented concurrently [11,12]. Two important characteristics make this effect unique. First, the change induced in perception is categorical rather than gradual. Only one other categorical multisensory illusion is known thus far, namely the McGurk effect. Secondly, audition dominates vision, which is rare. More often, as in the McGurk effect, our primary sense, vision, dominates audition.

In their experiments, Shams et al. paired a brief flash with 1-4 beeps. The most remarkable finding was that, on average, approximately 2 flashes were perceived when 1 flash was paired with 2 beeps. The number of perceived flashes was not significantly different from that of the control condition where 2 actual flashes were presented alone. Shams et al. concluded that the second beep caused subjects to perceive an illusory flash. This perceptual *fission* of a single flash due to multiple beeps was *not* matched by a perceptual *fusion* of multiple flashes due to a single beep. When 1-4 flashes were presented with 1 beep, visual perception was not different from when 1-4 flashes were presented alone.

On the basis of their results, Shams et al. suggested the *discontinuity hypothesis* stating that the influence of the discontinuous modality on the continuous modality is stronger than vice versa. They also noticed that the *modality appropriateness hypothesis* does not explain their results. According to this hypothesis, the dominant modality in their experiment would be audition because audition is superior to vision in temporal discrimination [16]. But this would not explain why only fissions and not fusions occurred.

It is important to distinguish between necessary and sufficient conditions when discussing these hypotheses. Whereas Shams et al.'s results indicate that discontinuity is a necessary condition for dominance, they do not address whether it is a sufficient condition. If it is, then it would be possible to reverse the direction of the illusion. By instructing subjects to count beeps and presenting more flashes than beeps, an auditory illusion of an additional beep should be created. In the same line of thought, the results indicate that modality appropriateness is not a sufficient condition for modality dominance, since then we would expect both fissions and fusions, but do not address whether it is a necessary condition. Only the existence of illusory beeps would falsify modality appropriateness as a necessary condition. Therefore, it is important to test whether this corresponding auditory phenomenon exists.

Challenging the modality appropriateness hypothesis is the *information reliability hypothesis* that suggests that the dominant modality is not defined by its appropriateness per se, but rather by how reliable it is. In addition to the appropriateness of the modality, at least two other factors may partake in determining its reliability [10]. First, cognitive factors may affect the reliability as when subjects are manipulated to believe one modality to be more reliable than another [15]. Second, stimulus properties affecting discrimination often affect the influence of the modality. As an example, take audiovisual speech perception in noise, where the discriminability of auditory speech decreases with increasing noise levels causing subjects' responses to approach randomness. However, if auditory speech is paired with a conflicting visual utterance, the randomness of the response distribution does not increase with increasing noise. Rather, it approaches that of the visual stimulus when presented alone [2]. Somehow, the randomness, or unreliability, of the auditory modality is discarded. Hence the integrated percept has favoured the more reliable visual modality. To assure that we do not accept the appropriate modality hypothesis when it is in fact information reliability that governs the cross-modal interactions, it is necessary to ascertain that the role of the dominant and dominated modalities does not reverse at any point throughout the range of information reliability.

The notion of reverting the illusion so as to produce visually induced illusory beeps only makes sense if the two modalities are not completely fused. In the case of complete fusion, one cannot tell flashes from beeps and therefore task instructions should have no effect on responses. On the other hand, if audiovisual perception allows access to the auditory and visual percepts separately, response patterns should be different when subjects are instructed to count beeps, as opposed to counting flashes. Warren named the possible effect of task instructions the *directed attention hypothesis* [15] because different response patterns are created by directing subjects' attention to either audition or vision. Task instructions might have no effect also for another reason: One modality might attract attention automatically, especially if stimuli in that modality are far more salient than those of the other modality, as happens when we are unable to read in a noisy environment. In any case, if the directed attention hypothesis does not hold, counting flashes or beeps should yield the same response for the same audiovisual stimulus.

We have outlined four hypotheses each advocating an attribute putatively characterizing the dominant modality: discontinuity, modality appropriateness, information reliability and directed attention. The purpose of the current experiments was to test how these theories apply to audiovisual integration of rapid flashes and beeps as evidenced by Shams et al.'s illusion. We employed an extended factorial experimental design presenting 1-3 beeps, 1-3 flashes and the 9 possible audiovisual combinations. We further divided our experiments in two blocks: one in which subjects were instructed to count the flashes and another in which they were instructed to count the beeps. This yields specific predictions from three of four of the mentioned hypotheses. Discontinuity as a sufficient condition for modality dominance predicts the existence of illusory flashes when counting flashes accompanied by more beeps than flashes. It also predicts illusory beeps when counting beeps accompanied by more flashes than beeps. In contrast, modality appropriateness as a necessary condition predicts that flashes cannot cause illusory beeps. Directed attention as a necessary condition predicts a difference in perception depending on whether one counts flashes or beeps. To test the information reliability hypothesis, we manipulated intensity of the sound stimuli. In our first experiment, it was at a clearly audible level of 80 dB(A) as was that employed by Shams et al. In our second experiment, it was near subjects' hearing threshold at 10 dB(A). The information reliability hypothesis predicts that the visual illusion would be weaker at a low sound level. More interestingly, it also predicts an auditory illusion at this low sound level.

2. Methods

2.1 Subjects

10 subjects (5 females, mean age 19 years) participated in Experiment 1. 9 subjects (3 females, mean age 23 years) participated in Experiment 2. All subjects were naïve as to the purpose of the experiment. Each subject's auditory threshold corresponding to the stimulus tone of 3.5 kHz was estimated by averaging thresholds for 3 and 4 kHz tones in left and right ear separately obtained using a Voyager 522 audiometer from Madsen Electronics. No subjects had an auditory threshold lower than 0 dB. However, in Experiment 2, 4 subjects did not show any influence of the number of beeps in the unimodal auditory condition. ($p>0.006$ Fisher's exact 3-by-3 test, significance level Bonferroni corrected for 9 independent tests). We conclude that the beeps were below their auditory threshold. This is likely to be due to the short duration of the beeps in comparison with the longer beeps used by the Voyager 522 audiometer. The 4 subjects were excluded from further analysis.

2.2 Stimuli

The stimuli were similar to those used in Shams et al.'s original experiment. Unimodal visual stimuli were 1, 2 or 3 flashes. A flash was a white disk with luminance 148 cd/m^2 on a black background with luminance 2.07 cd/m^2. The duration of a flash was 16.7 ms. The radius of the disk was 2° and the center of the disk was 5° below a fixation cross. Unimodal auditory stimuli were 1, 2 or 3 beeps. A beep was a Hamming windowed sine-wave with a frequency of 3.5 kHz and duration of 7.0 ms. The audiovisual stimuli were the 9 combinations of the 3 auditory and the 3 visual stimuli. Stimulus onset asynchrony

(SOA) was 67 ms for both auditory and visual stimuli. Since Shams et al.'s results indicated that the effect is largest when the audiovisual asynchrony is small, the first beep and flash were delivered simultaneously with a maximal asynchrony of 1 ms. The only difference in stimulus attributes between Experiment 1 and 2 was the sound level. In Experiment 1, the sound level was 80 dB(A), a clearly audible level as in Shams et al.'s experiment. In Experiment 2, the sound level was 10 dB(A), close to subjects' threshold. The sound level in Experiment 2 was chosen on the basis of pilot studies (data not shown) which indicated little effect of decreasing sound level until 10 dB(A) was reached.

2.3 Procedure

Both experiments consisted of two blocks. In one block, subjects were instructed to count the flashes and in the other they were instructed to count the beeps. The count-flashes block contained all 3 unimodal visual and all 9 audiovisual stimuli. The count-beeps block contained all 3 unimodal auditory and all 9 audiovisual stimuli. The order of the blocks was varied pseudo-randomly between subjects. All stimuli in a block were presented 20 times in pseudo-random order.

Subjects sat in a sound attenuated room with normal ambient lighting. Visual stimuli were presented on a monitor 110 cm in front of the subject. Auditory stimuli were presented through headphones. Before each block, subjects were given written instructions to maintain their gaze at the fixation cross and to count either flashes or beeps.

3 Results

3.1 Experiment 1

3.1.1 Count-flashes block

The mean response rates across subjects from the count-flashes block are plotted in Figure 1. Graphs in the first column, embedded in a dark gray background, display the results of the unimodal visual conditions. Subjects identified 1 and 2 flashes well but 3 flashes were often confused with 2 flashes. This is evidenced by correct identification rates of 82%, 71% and 50% for 1-3 flashes respectively.

We divide illusory effects into fissions, in which more beeps than flashes cause the perception of more flashes than actually presented, and fusions, in which fewer beeps than flashes cause the perception of fewer flashes than actually presented. Three stimuli could give rise to fissions: 1 flash with 2 or 3 beeps; and 2 flashes with 3 beeps. The results are depicted in graphs embedded in a light gray background in Figure 1. The effect of the beeps is obvious when these graphs are compared with those in the unimodal visual condition. Overall, subjects' visual percepts were strongly influenced by the number of concurrent beeps. For 1 flash with 2 beeps subjects perceived 2 flashes in 55% of all trials. When 1 flash was paired with 3 beeps, subjects perceived 3 flashes in 24% of all trials. With 2 flashes with 3 beeps, subjects perceived 3 flashes in 45% of all trials.

To test the significance of fission effects, we formed a fission response count by pooling the response count across response categories indicating more flashes than actually presented. We then formed a no-fission response count by pooling response counts across response categories indicating the correct or fewer than correct number of flashes. The ratio of these two pooled counts is the odds of fissions versus no fissions. For each fission stimulus, we compared these odds with the odds in the corresponding unimodal condition thus forming the odds ratio. The odds ratio thus reflects the magnitude of the effect, an odds ratio greater than 1 indicating an increase in fissions in the audiovisual case as compared to the unimodal case. We tested the significance of the odds ratio being greater than 1 using Fisher's exact, one-sided test. When response counts were pooled across subjects, all three possible fission stimulus sets gave highly significant fission responses ($p<10^{-15}$ for each stimulus set).

Three stimuli could give rise to fusions: 2 or 3 flashes with 1 beep; and 3 flashes with 2 beeps. Graphs describing these conditions are embedded in a medium grey background in Figure 1. Also here the number of concurrently presented beeps influenced visual perception. When 2 flashes were paired with 1 beep, subjects perceived 1 flash in 57% of all trials. When 3 flashes were paired with 1 beep, subjects perceived 1 flash in 22% of all trials and when 3 flashes were paired with 2 beeps subjects perceived 2 flashes in 64% of all trials. The statistical analysis of subjects' responses to fusion stimuli is analogous to that of fission stimuli except we here compare fusion (fewer than correct) counts with no fusion counts (correct or more). When response counts were pooled across subjects, fusion effects were significant with all three stimulus sets ($p<10^{-3}$ for every stimulus set).

To investigate the effect of individual variation we pooled response counts across the three fission stimuli and not across individuals. In 8 of 10 subjects fission effects were significant at the Bonferroni corrected level of 0.005. The corresponding analysis for fusion counts showed significant fusion effects in 6 of 10 subjects. Two of the subjects did not show either fission or fusion effects. This indicates that the phenomenon is generally, but not ubiquitously present in the population.

The clear presence of both fissions and fusions contradicts the hypothesis that discontinuity is a necessary condition for the dominant modality. A weaker hypothesis, as proposed by Shams et al., would be that discontinuous stimuli have a greater effect on continuous stimuli than vice versa. In that case, the odds ratio for fissions should be greater than that for fusions. We tested this by pooling response counts across both all subjects and either fission or fusion stimuli. The corresponding contingency tables are given in Table 1. The fission odds ratio was greater than that of the fusions. Woolf's [1,5] test confirmed this by rejecting homogeneity of the odds ratios (p=10^{-4}).

3.1.2 Count-beeps block

The mean response rates across subjects from the count-beeps block are plotted in Figure 2. The graphs embedded in a medium gray background in the first row depict the results of the auditory only condition. All subjects performed well in discriminating between 1, 2 and 3 beeps as evidenced by a mean correct identification rates of 97%, 89% and 93%, respectively. The subsequent rows of graphs depict the results when 1-3 flashes were presented concurrently with the beeps and are almost identical to the unimodal auditory graphs. There was clearly no effect of concurrent presentation of flashes (p>0.38, for any subject and any number of beeps, Fisher's exact 4x3 test).

3.2 Experiment 2

3.2.1 Count-flashes block

Figure 3 shows the mean response rates across the 5 subjects who showed influence from auditory stimulation. As in Figure 1, the graphs in the leftmost column display the results from the unimodal visual trials where the correct identification rates were 92%, 70% and 74% for 1-3 flashes respectively. We expected the response rates to be similar to those in Experiment 1 since the stimuli in the two experiments were identical. A comparison of Figure 3 with Figure 1 reveals, however, that the number of correct identifications was significantly higher in Experiment 2 than in Experiment 1 for 1 flash ($p<0.01$), and 3 flashes ($p< 10^{-4}$), but not for 2 flashes ($p>0.2$, Fisher's exact one-sided 2-by-2 test). This was not due to the exclusion of 4 subjects who had lower auditory – and thus possibly also visual – sensitivity since the effect persisted when all 9 subjects were included in the analysis. We can offer no explanation as to why 1 and 3 flashes were more accurately identified in Experiment 2, except random fluctuations. However, all our statistical analyses of audiovisual perception are based on the change *relative* to unimodal perception. Therefore, this effect should have little consequence as to our general conclusions.

The analysis of fission and fusion effects was done as in Experiment 1. When response counts were pooled across subjects, all three possible fission stimuli gave significant fission responses (p<0.0002). In contrast, none of the fusion stimuli caused significant fusion responses (p>0.31 for each stimulus). When response counts were pooled across the three fusion stimuli, fusion responses were still not significant (p>0.36). We conclude that whereas visual fissions persist even when auditory stimulation is near threshold, visual fusions are eliminated at this extreme. The contingency tables for fission and fusion counts pooled across both stimuli and subjects are given in Table 2.

As expected, the change of sound level affected the magnitude of the illusion. Comparing the contingency tables from Table 2 with those from Table 1, we see that the odds ratios were higher for both fissions (p>0.3, Woolf's test) and fusions (p=0.001, Woolf's test) in Experiment 1 than in Experiment 2. Although the effect was not significant for fissions, we conclude that the magnitude of the visual illusion tends to decrease with the sound level.

3.2.2 Count-beeps block

With the sound level near auditory threshold, this condition tests whether very low auditory information reliability allows vision to be the dominant modality. The average response rates across 5 subjects are plotted in Figure 4. As in Figure 2, the top row of graphs depicts the results from the unimodal auditory condition where the correct identification rates were 92%, 61% and 55% for 1-3 beeps respectively.

In contrast to the count-beeps block in Experiment 1, the response rate was highly dependent on the number of flashes presented concurrently with the beeps. This is seen in the variation across columns in Figure 4. Each of the three fission stimuli, 1 beep with 2 or 3 flashes, and 2 beeps with 3 flashes, had a significant fission effect ($p<0.003$ for each stimulus).

Only one of the fusion stimuli, 3 beeps with 1 flash, showed a significant fusion effect ($p<10^{-12}$). When response counts were pooled across all three fusion stimulus sets, the effect was highly significant ($p<0.002$). Table 3 gives the contingency tables and odds ratios with response counts pooled across subjects, and fission or fusion response counts. As in the count-flashes block in Experiment 1, the difference in odds ratios from fission and fusion effects was significant ($p=3\times10^{-4}$, Woolf's test) indicating that the fission effect was also here stronger than the fusion effect.

The hypothesis of directed attention predicts a difference in subjects' responses based on whether they count flashes or beeps. In Experiment 1, this effect was very strong with no visual influence on the number of perceived beeps as opposed to strong auditory influence on the number of perceived flashes. Also in Experiment 2 the count-beeps block and the count-flashes block showed significant differences for each incongruent stimulus combination ($p<0.03$, Fisher's exact, two-sided 2-by-3 test).

4. Discussion

In the introduction we outlined four hypotheses on what determines the dominating modality: discontinuity, modality appropriateness, information reliability and directed attention. We now discuss how our results and analyses pertain to each of them.

Our most surprising result relates to the discontinuity hypothesis. We found a strong visual fusion illusion — much stronger than that reported by Shams et al. This effect was highly significant, whereas Shams et al. concluded that it was negligible [14]. We set out to determine whether discontinuity is a sufficient condition for the illusion to occur but found that it is not even a necessary condition. We did, however, find that that fusion effects — auditory as well as visual — were consistently weaker than fission effects, in agreement with Shams et al. Hence our results confirm Shams and coworker's weaker form of the discontinuity hypothesis: Stimulus discontinuity is a factor which increases the tendency of a modality to dominate.

Our finding of visually induced illusory beeps when the sound level was low refutes modality appropriateness as a necessary condition and supports the information reliability hypothesis. Information reliability worked as a factor that influenced the tendency of a modality to dominate. This is seen also by comparing the strength of the visual illusion in Experiments 1 and 2. We found illusory effects in both cases but in Experiment 2, where audition was less reliable, they were weaker.

However, there is a caveat to our rebuttal of the modality appropriateness hypothesis. Our pilot tests (data not shown here) indicated that it was indeed necessary to decrease the sound level to near auditory threshold in order to elicit the auditory illusion. This is not the case for the visual illusion which readily occurs when the visibility of the flashes is far above threshold even if the beeps are near threshold as in Experiment 2. Hence beeps could cause illusory flashes throughout a greater range of relative information reliability.

This asymmetry could be due to audition being more appropriate for the task of counting rapid events.

Information reliability might also provide an explanation as to why fusion effects were significant in our experiments but not in those of Shams et al. In Experiment 2, the decrease in sound level as compared to Experiment 1 decreased fission effects while eliminating fusion effects. If some confounding factor in the experimental setup made illusory effects weaker in the experiments of Shams et al. than they were in ours, it could have merely weakened the fission effect while eliminating the weaker fusion effect. But judging from the results of Shams et al., it seems that the fission effect was, in fact, stronger than in our experiment.

Task instructions had a strong influence on subjects' responses. Most obvious was the complete absence of visual influence on counted beeps in Experiment 1 compared with strong influence from both vision and audition on counted flashes. Response patterns differed significantly between count-beeps and count-flashes blocks also in Experiment 2. These results provide support for the directed attention hypothesis, indicating that the incongruent audiovisual stimuli did not necessarily produce a compulsorily unified percept.

Salient stimuli automatically attracting attention could not be the sole cause of the visual illusion. This is shown by the count-flashes block in Experiment 2. Here, the beeps were much less salient and hence unlikely to distract attention involuntarily from the flashes, but still the visual illusion occurred. However, we cannot exclude the possibility that exogenous attention had an effect in the count-beeps block. This could have inflated the magnitude of the perceptual effect and even be the main cause of the illusory beeps, which would then be a response bias rather than a true perceptual effect.

In conclusion, our results not only confirm the robustness of the illusion discovered by Shams and coworkers but also reveal a corresponding fusion illusion. We have also discovered an equivalent visually induced auditory illusion, which may, however, be caused by a response bias. Our theoretical considerations on the discontinuity, modality appropriateness and information reliability hypotheses suggest that all of these should be considered as factors which contribute to the relative dominance of each modality and not as all-or-nothing conditions. Finally, the clear effect of task instructions indicates that rapid flashes and beeps are not automatically integrated to a unified percept.

Acknowledgements

The authors thank Ms. Reetta Korhonen for practical assistance and an anonymous reviewer for suggesting improvements of the original manuscript. The study was financially supported by the Academy of Finland and the MUHCI European Union Research Training Network.

References

[1] Agresti, A., *Categorical Data Analysis*, John Wiley, 2002.

[2] Andersen, T.S., Tiippana, K., Lampinen, J. and Sams, M., Modelling of Audiovisual Speech Perception in Noise. *Proceedings of the Fourth International ESCA ETRW Conference on Auditory-Visual Speech Processing*, Ålborg, Denmark., 2001, pp. 172-176.

[3] Benoît, C., Mohamadi, T. and Kandel, S., Effects of Phonetic Context on Audio-Visual Intelligibility of French, *Journal of Speech and Hearing Research*, 37 (1994) 1195-1203.

[4] Green, K., Kuhl, P., Meltzoff, A. and Stevens, E., Integrating speech information across talkers, gender, and sensory modality: female faces and male voices in the McGurk effect., *Perception & Psychophysics*, 50 (1991) 524-36.

[5] Lumley, T., The rmeta Package. http://cran.r-project.org/src/contrib/PACKAGES.html#rmeta, 2003.

[6] MacDonald, J. and McGurk, H., Visual influences on speech perception processes, *Perception & Psychophysics*, 24 (1978) 253-257.

[7] MacLeod, A. and Summerfield, Q., Quantifying the contribution of vision to speech perception in noise, *British Journal of Audiology*, 21 (1987) 131-41.

[8] McGurk, H. and MacDonald, J., Hearing lips and seeing voices, *Nature*, 264 (1976) 746-748.

[9] Sams, M., Manninen, P., Surakka, V., Helin, P. and Kättö, R., McGurk effect in Finnish syllables, isolated words, and words in sentences: effects of word meaning and sentence context, *Speech Communication*, 26 (1998) 75-87.

[10] Schwartz, J.-L., Robert-Ribes, J. and Escudier, P., Ten years after Summerfield: A taxonomy of models for audio-visual fusion in speech perception. In R. Campbell (Ed.), *Hearing by Eye: The Psychology of Lipreading*, Lawrence Earlbaum Associates Ltd., Hove, U.K., 1998, pp. 3-51.

[11] Shams, L., Kamitani, Y. and Shimojo, S., Illusions. What you see is what you hear, *Nature*, 408 (2000) 788.

[12] Shams, L., Yukiyasu, K. and Shimojo, S., Visual illusion induced by sound, *Cognitive Brain Research*, 14 (2002) 147-152.

[13] Stein, B.E. and Meredith, M.A., *The Merging of the Senses*, A Bradford Book, Cambridge, MA, 1993, 211 pp.

[14] Sumby, W. and Pollack, I., Visual contribution to speech intelligibility in noise, *Journal of the Acoustical Society of America*, 26 (1954) 212-215.

[15] Warren, D.H., Spatial localization under conflict conditions: is there a single explanation?, *Perception*, 8 (1979) 323-37.

[16] Welch, R.B. and Warren, D.H., Immediate Perceptual Response to Intersensory Discrepancy, *Psychological Bulletin*, 88 (1980) 638-667.

Figures

1

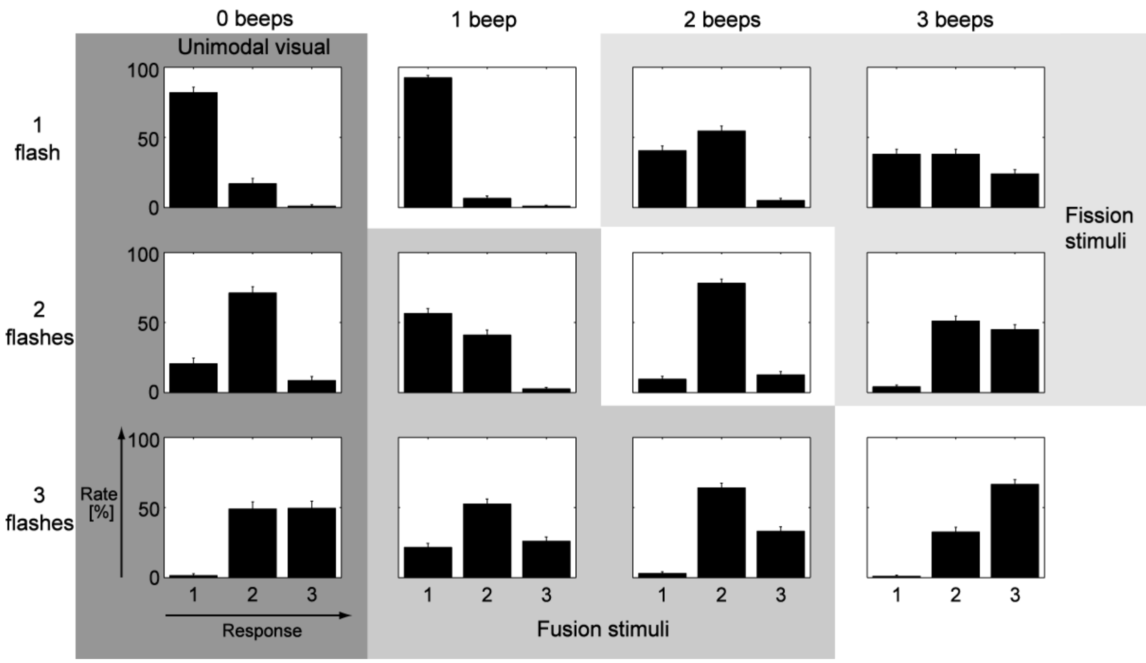

Figure 1. Mean response rate for counted flashes at 80 dB sound intensity. Graphs in columns 1-4 displays data from trials employing 0-3 beeps. Hence, first column, embedded in dark gray, corresponds to unimodal visual stimuli. Rows 1-3 display data from trials employing 1-3 flashes. Fission stimuli are framed in a light gray box and fusion stimuli in a medium gray box. Errorbars indicate standard error of the mean (SEM).

2

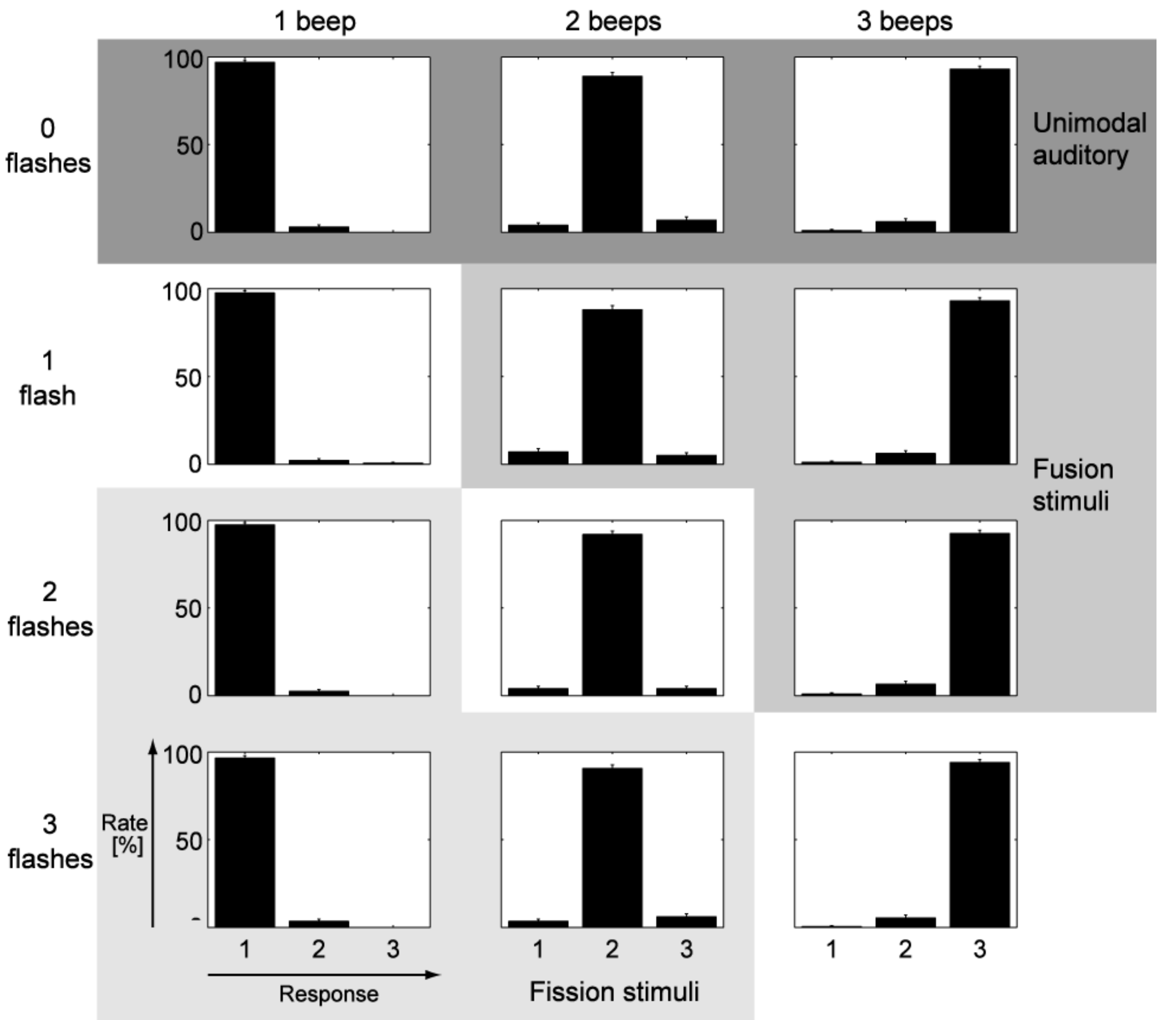

Figure 2 Mean response rate for counted beeps at 80 dB sound intensity. Graphs in rows 1-4 display data from trials employing 0-3 flashes. Hence, first row, embedded in dark gray, corresponds to unimodal auditory stimuli. Columns 1-3 display data from trials employing 1-3 beeps. Fission stimuli are framed in a light gray box and fusion stimuli in a medium gray box. Errorbars indicate standard error of the mean (SEM).

3

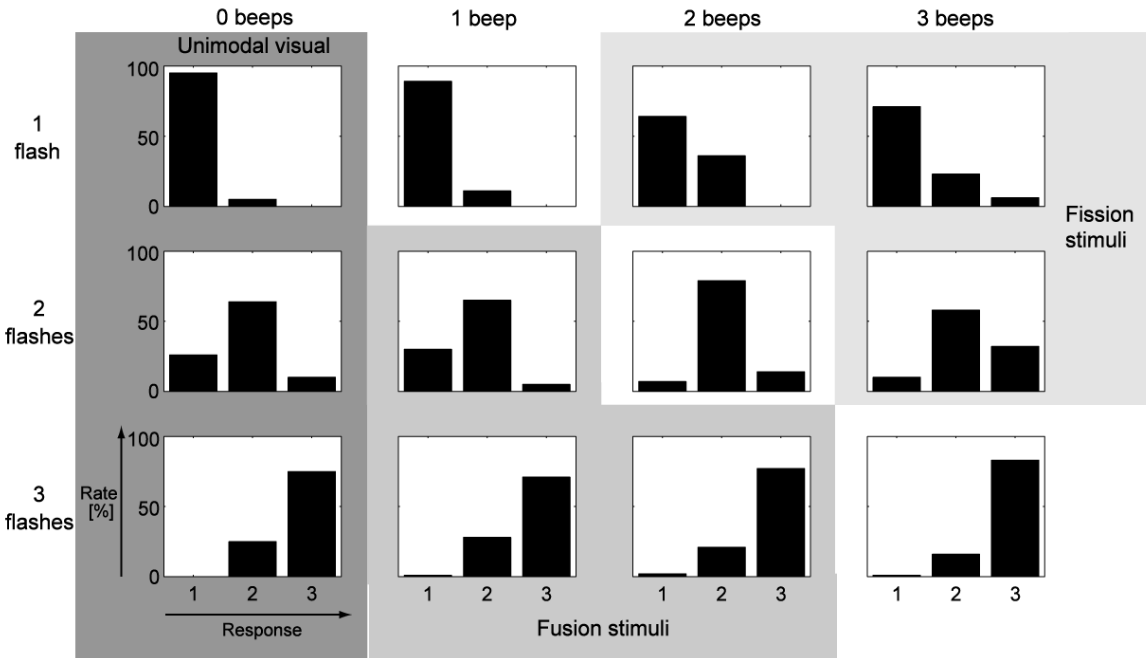

Figure 3 Mean response rate for counted flashes at 10 dB sound intensity. Details as in Fig. 1.

4

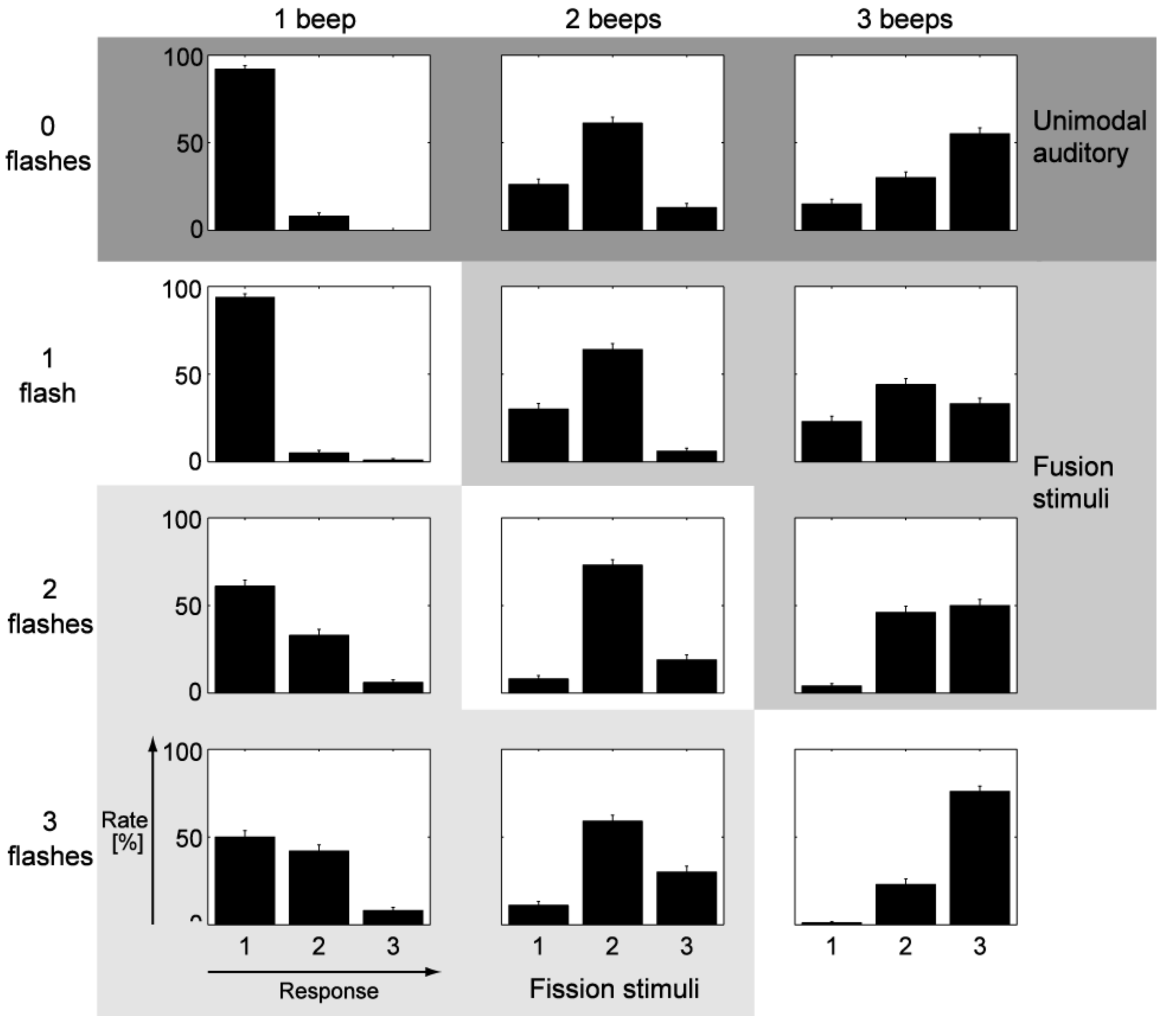

Figure 4 Mean response rate for counted beeps at 10 dB sound intensity. Details as in Fig. 2.

Tables

1

	Fissions	No fission	Sum	Fusions	No fusion	Sum
Audiovisual	333	267	600	395	205	600
Visual	53	347	400	142	258	400
Odds ratio	8.17			3.50		

Table 1. Contingency tables with odds ratios for visual fission (left) and fission (right) effects pooled across stimuli and subjects in Experiment 1.

2

	Fissions	No fission	Sum	Fusions	No fusion	Sum
Audiovisual	97	203	300	82	218	300
Visual	15	185	200	51	149	200
Odds ratio	5.89			1.10		

Table 2 . Contingency tables with odds ratios for visual fission (left) and fusion (right) effects pooled across stimuli and subjects in Experiment 2.

3

	Fissions	No fission	Sum	Fusions	No fusion	Sum
Audiovisual	119	181	300	147	153	300
Visual	21	179	200	71	129	200
Odds ratio	5.60			1.75		

Table 3. Contingency tables with odds ratios for auditory fission (left) and fusion (right) effects pooled across stimuli and subjects in Experiment 2.