A PSYCHOACOUSTIC STUDY OF THE MECHANISM OF BINAURAL FUSION

by

JOSEPH LINDLEY HALL, II

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May, 1959

Signature of Author

Department of Electrical Engineering, May 27, 1959

Certified by ____________ Thesis Supervisor

Accepted by ____________ Chairman, Departmental Committee on Graduate Students
A PSYCHOACOUSTIC STUDY OF THE MECHANISM OF BIN AURAL FUSION

by

JOSEPH LINDLEY HALL, II

Submitted to the Department of Electrical Engineering on May 1959 in partial fulfillment of the requirements for the degree of Master of Science

ABSTRACT

A series of psychoacoustic experiments was performed in order to determine some of the processes involved in binaural fusion. The experiments consisted of investigating under what conditions a human listener could resolve stimuli presented at the two ears into a single virtual image at the center of his head. The stimuli used, clicks and noise bursts, had previously been investigated neurophysiologically.

The experiments performed were: 1) With a train of clicks in one ear and a train of noise bursts in the other (signals presented at the same repetition rate), the effect of masking noise on either or both stimuli was studied. 2) With a train of noise bursts in one ear and a train of noise bursts in the other (signals of the same duration and presented at the same repetition rate, but uncorrelated), the effect of masking noise on either or both stimuli was studied. 3) With a train of single clicks in one ear and a train of pairs of closely spaced clicks in the other (signals presented at the same repetition rate), the effect of masking noise on the train of pairs of clicks was studied. 4) With a train of clicks in one ear and a train of noise bursts in the other (signals presented at the same repetition rate), the effect of changing the noise burst duration was studied.

The following model of the mechanism of binaural fusion is proposed: The stimulus arriving at the ear produces a timing signal. The timing signals from the two ears travel to a centrally located fusion mechanism. Depending on the nature of the timing signals from the two ears, the model predicts whether the subject will judge that he hears a fused virtual image or two separate sources.
The strength of the timing signal is determined by the short-term integrated neural response to a stimulus. The timing signal can be weakened by reducing the number of neural units firing in response to a stimulus or by desynchronizing their times of firing.

The fusion mechanism operates by comparing the strength of the timing signals from the two sides. An increase in strength of one timing signal can be compensated for by making that timing signal arrive later in time.

The fusion mechanism is able to tolerate only a certain discrepancy in the strength of the timing signals from the two sides. If this discrepancy is too great, fusion is not predicted by the model. In addition, fusion is not predicted if the timing signals become too weak.

The experimental work was done at the Murray Hill Laboratories of Bell Telephone Laboratories.
ACKNOWLEDGMENT

I wish to thank Professor Moise H. Goldstein, Jr., for encouraging me to undertake this program, and for his valuable criticism of the manuscript. I am indebted to Dr. Edward E. David, Jr., and to Dr. Willem A. van Bergeijk for many helpful suggestions while the experimental work was in progress. I am particularly indebted to Dr. van Bergeijk for assistance in the preparation of a report on this work for Bell Telephone Laboratories. I also wish to express my gratitude to the patient subjects who made this work possible.
# TABLE OF CONTENTS

1. ABSTRACT. ...................................................... 11
2. ACKNOWLEDGMENT. .............................................. IV
3. TABLE OF CONTENTS .............................................. V
4. LIST OF FIGURES AND TABLES. ................................. VII
5. INTRODUCTION. .................................................. 1
6. MODELS AND DATA RELATED TO THE PROBLEM. .............. 5
7. PURPOSE OF THE PRESENT STUDY. ............................. 10
8. EXPERIMENTS ................................................... 11
   8.1 Procedure Common to All Experiments. ................... 11
   8.2 Experiment I. Effect of Masking Noise on Fusion of a Train of Clicks with a Train of Noise Bursts. .......... 15
   8.2.1 Experimental Procedure ................................. 15
   8.2.2 Results .................................................. 19
   8.3 Experiment II. Effect of Masking Noise on Fusion of Two Trains of Noise Bursts .................................. 25
   8.3.1 Experimental Procedure ................................. 25
   8.3.2 Results .................................................. 26
   8.4 Experiment III. Effect of Masking Noise on Fusion of a Train of Single Clicks with a Train of Closely Spaced Pairs of Clicks. ........................................ 26
   8.4.1 Experimental Procedure ................................. 26
   8.4.2 Results .................................................. 29
4.5  Experiment IV. Effect of Noise Burst Duration On Fusion of a Train of Noise Bursts with a Train of Clicks ........... 31
4.5.1  Experimental Procedure. ................... 31
4.5.2  Results .................................... 34

V.  INTERPRETATION AND DISCUSSION. ................. 37
5.1  Proposed Model. ............................... 37
5.2  Interpretation of Experimental Results
     In Terms of Proposed Model. .................. 38
5.2.1  Experiment I. ............................. 38
5.2.2  Experiment II .............................. 42
5.2.3  Experiment III. ......................... 43
5.2.4  Experiment IV ............................ 45
5.3  Alternate Models ............................ 47

VI. CONCLUSIONS. ................................. 48

BIBLIOGRAPHY ..................................... 50
LIST OF FIGURES AND TABLES

Figure 1. Heuristic representation of neural unit firing probability for the following stimuli: (a) Click. (b) Noise burst. (c) Click in the presence of masking noise. (d) Noise burst in the presence of masking noise. 8

Figure 2. Schematic representation of laboratory setup. Noise sources uncorrelated. 12

Figure 3. Frequency response of headphones. 14

Figure 4. Schematic representation of stimuli used in experiments 16

Figure 5. Noise burst gate. (a) Multivibrator used to drive relay. (b) Contact arrangement on relay. 18

Figure 6. Experiment I. Train of clicks in left ear, train of noise bursts in right ear. Effect of masking noise on $\tau_{RL}$. 20, 21

Figure 7. Experiment I. Train of clicks in left ear, train of noise bursts in right ear. Effect of masking noise in preventing fusion. 22

Figure 8. Experiment II. Train of noise bursts in left ear, train of noise bursts in right ear. Effect of masking noise in preventing fusion 27
Figure 9. Experiment III. Train of single clicks in left ear, train of pairs of clicks in right ear. Proportion of trials in which subject is able to get two centering positions of single click. 30

Figure 10. Electronic noise burst gate. 33

Figure 11. Experiment IV. Train of clicks in left ear, train of noise bursts in right ear. Effect of burst length on interval between click and noise burst for a centered image. 35

TABLES

Table I. Experiment I. Train of clicks in left ear, train of noise bursts in right ear. 24

Table II. Experiment II. Train of noise bursts in left ear, train of noise bursts in right ear. 28

Table III. Experiment III. Train of single clicks in right ear, train of pairs of clicks in left ear. Interval at which subject is able to get two centering positions of single click in 50% of trials. 31
I. INTRODUCTION

In some way the nervous system acts upon acoustic stimuli so that when the sounds reaching the two ears are "similar", the listener hears a single fused source. The apparent location of this source depends on the relative intensities and relative time of arrival of the sounds at the two ears. This is the familiar process by which we can tell where a sound source is located in the absence of any visual cue. When the sounds reaching the two ears are different enough, either in time of arrival, intensity, or frequency content, this fusion dissolves and the listener hears two sources, one at each ear.

Very little is known about the method of operation of the nervous system in producing the impression of a single fused image. There is a large body of literature concerned with binaural phenomena, but the majority of it takes a phenomenological approach, with little consideration of the physiological processes involved. Furthermore, attempts to develop a physiological model have been handicapped by the extreme complexity of the nervous system.

Békésy\(^1\) carried out extensive investigation of the fusion process. His experiments have included measuring the amplitude difference and the time difference between stimuli at the two ears necessary to
move the apparent source to the side of the head. He hypothesized that fusion is governed by a central cell body, which is connected to the outputs of both ears. Individual cells in this central body are "tuned" to one side of the head or the other, depending on the intensity and time of arrival of the signals from the two sides. The apparent location of the source is determined by the relative number of cells tuned in the two directions.

Jeffress, et al., 2, 11, 20, 30 investigated, among other things, the minimum discernible interaural time delay and the accuracy with which a listener locates sounds of different sorts. The stimuli used were sinusoids and narrow bands of noise. These experiments supported the commonly accepted observation that time difference plays the predominant role in determining apparent location of a source at low frequencies, while at high frequencies, intensity differences are more important. Jeffress attributed this effect to variable physiological delays. He developed a model 20 of the fusion process in which the central fusion mechanism compares activity in the corresponding portions of the cochleas at the two sides.

In more recent work, Deatherage and Hirsch 10 investigated the effect of intensity differences and of the presence of high-frequency masking noise on the
fusion of two clicks. Combining their results with neurophysiological work performed by Deatherage, Elridge, and Davis, they concluded that the fusion process operates by comparing the relative arrival time of synchronous neural volleys (the $N_1$, $N_2$ configuration), from the two sides of the head, and that these synchronous volleys arise primarily in the basal turn of the cochlea.

Sayers and Cherry have developed an elaborate model of the binaural fusion process. The model is based on data obtained by presenting a listener with sine waves, intoned vowels, or running speech at his two ears, and forcing a judgment of "source left of center" or "source right of center". Essentially, the model operates by taking the short-term cross-correlation of the signal plus average signal level at the two ears and time averaging the result. The model has little physiological fundation.

We have conducted a series of psychoacoustic experiments in binaural fusion, using stimuli which had previously been studied neurophysiologically. Our experimental results are interpreted in terms of a heuristic model of neurophysiological events.

---

*heuristic: "Serving to . . . stimulate (further) investigation", Webster's New Collegiate Dictionary, 1949."
There are two reservations we should bear in mind: 1) In a free-field condition, the single fused source appears to be located somewhere outside the listener. If the sounds are being administered through earphones, the listener judges the source to be located inside his head. Our studies concern the latter phenomenon. While the reason for this dichotomy is not fully understood, it is commonly assumed that these two types of fusion are mediated by the same mechanism of binaural fusion. 2) The fusion mechanism referred to in this paper is a heuristic model of a neurophysiological process. It is not the sort of thing that could be removed surgically from an experimental animal and studied in isolation.
II. MODELS AND DATA RELATED TO THE PROBLEM

David, Guttman, and van Bergeijk\textsuperscript{6,7} and Deatherage and Hirsch\textsuperscript{10} have shown in a series of psychoacoustic experiments that, for impulsive stimuli, time and intensity can be "traded" in fusion. If trains of electric pulses of 100-\textmu sec duration (clicks) are fed to the subject through earphones, the apparent source of the clicks can be moved to the left side of the head, either by making the click on the left more intense than the click on the right or by making the click on the left precede the click on the right. If the click on the left is more intense than the click on the right, the apparent source can be moved back to the center of the head by making the click on the right arrive sooner than the click on the left.

Goldstein and Kiang\textsuperscript{17,18} have conducted a series of neurophysiological experiments investigating monaural phenomena. These experiments were done with gross electrodes placed at the round window and in the auditory cortex of cats. When the stimulus was a click, if enough masking noise was added to eradicate the characteristic neural response to the click at the round window (the $N_1$, $N_2$ peak configuration), the neural response at the auditory cortex would also be eradicated. When the stimulus was the sudden onset of
wide-band gaussian noise (the beginning of a noise burst), more masking noise was required to eradicate the neural response at the auditory cortex than to eradicate the $N_1, N_2$ peaks at the round window. Their explanation of this was that masking noise desynchronized the response to the noise burst but not to the click and that desynchronization destroyed the response at the round window but not at the auditory cortex.

This interpretation may be elucidated in terms of a simple mathematical model. Consider a "neural unit". This neural unit is a mathematical entity, and need have no physical significance. If we must pin it down to something in the nervous system, it would most closely resemble a neuron or group of neurons. Let $P(t)$ represent the instantaneous firing rate of a population of neurons (i.e., $P(t) \Delta t$ is the probability that a unit of the population fires in the very small interval of time $t \leq t' < t + \Delta t$). $P(t)$ depends on the stimulus intensity and on the history of the neural units in the population. If the stimulus is intense, the probability that a neural unit will fire is high. After many neural units fire, $P(t)$ decreases sharply, since many of the units are refractory, then $P(t)$ returns slowly to its original value.
We can give a heuristic picture of $P(t)$ for various stimuli (after Goldstein and Kiang\textsuperscript{18}). These $P(t)$ were not calculated from experimental data, but they do serve to explain the experimental results. Figure 1(a) shows $P(t)$ for a click. Initially, $P(t)$ is at some resting level, corresponding to the spontaneous discharge of neural units. $P(t)$ increases sharply at the occurrence of the click, then decreases below its resting level while units fired by the click are refractory. Eventually it returns to its resting level. Figure 1(b) shows $P(t)$ for a noise burst. Since very few neural units are refractory at the onset of the noise burst, $P(t)$ will show a sharp spike. While the units fired at the beginning of the noise burst are refractory, $P(t)$ will be low. As the thresholds of the units assume a steady state distribution, $P(t)$ will settle down at a level greater than the level in the absence of noise, but not as large as the initial peak. Figures 1(c) and 1(d) show the effect of masking noise on the responses to a click and to a noise burst. Masking noise added to the click produces an increase in base-line activity and a decrease in the height of the spike produced by the click. Masking noise added to the noise burst also increases the base-line activity and decreases the initial response to the noise burst, but since neural units which are refractory at the noise-burst
FIGURE 1

Heuristic representation of neural unit firing probability for the following stimuli: (a) Click. (b) Noise burst. (c) Click in the presence of masking noise. (d) Noise burst in the presence of masking noise.
onset recover and fire a short time later, the initial peak in \( P(t) \) is broader than it was when masking noise was absent. The addition of masking noise to a click reduces the number of units fired by the click, but it does not affect the times at which these units fire. Thus, since a click results in relatively synchronous firing of the responding units, a masking noise sufficient to mask the round window responses to clicks has suppressed the response to the click. However, when the stimulus is a burst of noise, the round window response disappears because of desynchronization of unit activity when there is still a response to the stimulus. This response can be seen at the cortex where the requirements of synchrony in responding units are much less stringent.
III. PURPOSE OF THE PRESENT STUDY

The Goldstein and Kiang experiments indicate that we have a way of controlling two parameters of the neural response to impulsive stimuli: we can desynchronize the times of firing of neural units by adding masking noise to a noise burst, or we can reduce the number of units firing without desynchronizing the response by adding masking noise to a click. This suggests that we can conduct a series of psychoacoustic experiments similar to those conducted by David, et al., but determine the importance of synchrony of neural discharges in binaural fusion by using as stimuli clicks and noise bursts in the presence of masking noise. The apparent trading relationship between time difference and intensity difference provides us with a qualitative measure of the "strength" of a signal for fusion, since we can measure time difference directly.
IV. EXPERIMENTS

4.1 PROCEDURE COMMON TO ALL EXPERIMENTS

The laboratory setup is illustrated in Figure 2. Each click generator is a one-shot multivibrator, which fires when a sawtooth common to all three click generators reaches a certain specified level. The repetition rate of the click generators can be controlled by adjusting the period of the sawtooth. For a given repetition rate, the timing can be controlled by adjusting the sawtooth level at which the generator fires. The output of each click generator is a 100-μsec rectangular pulse, which is the "click" stimulus, and a sync pulse, which can be used to trigger a noise burst. The sync pulse also triggers a Hewlett-Packard frequency counter, which measures the repetition rate and the interval between any two pulses.

The clicks or noise bursts are amplified, mixed with masking noise, and transmitted to the subject, who is sitting in a soundproof room. Using this apparatus, the experimenter can present the subject with any repetitive pattern of clicks or noise bursts, with up to three clicks or noise bursts per period. The subject can control the timing of one click generator by means of an unmarked control in the listening booth.

General Radio unit amplifiers were used throughout. Noise for the noise bursts was generated by
FIGURE 2
Schematic representation of laboratory setup. Noise sources uncorrelated.
General Radio random-noise generators. Masking noise was generated by Scott random-noise generators. All noise sources were uncorrelated. The subject wore a matched pair of PDR-8 earphones. Figure 3 shows the frequency response of the earphones, measured with a 6-cc coupler.

The subject went through a training and evaluation period before any formal data were recorded. Click trains of equal intensity at 20 db SL, 20 pps, were presented at each ear, and the subject was required to adjust the relative time of arrival of the two clicks for a centered image. DB SL corresponds to decibels with respect to sensation level. This is the level in db above threshold. We determined thresholds monaurally by the method of limits. Next a 10-msec noise burst at 20 db SL was substituted for one of the clicks, and the subject was required to get a centered image with masking noise added to the click and to the noise burst. This procedure gave the experimenter an idea of what sort of response could be expected from the subject. One subject was unable to fuse the click with the noise burst under any conditions. Another subject was very inconsistent in her responses. Both were rejected at this point. Out of five subjects (all with no previous psychoacoustic experience), three were usable. The three subjects used were male, ages 22 to 35 years.
FREQUENCY RESPONSE OF HEADPHONES
INPUT VOLTAGE = 0.1V RMS

FIGURE 3
We asked the subject to adjust the relative time of arrival of the stimuli until the apparent source was judged to be at the center of his head. Kikuchi\textsuperscript{21} has suggested what may be a more precise method of obtaining a centered image. The stimuli are reversed right and left periodically, and the subject adjusts the stimuli until the apparent source does not move when the stimuli change sides. Some of our results might have been quite different if we had used this method. The data may have been more consistent, since people seem to have trouble locating the center of their heads. Also, this method would have eliminated any bias produced by the presence of masking noise in one ear but not in the other. We did not use the method suggested by Kikuchi, as we wanted to be able to compare our results with those of David, \textit{et al.}

4.2. \textbf{EXPERIMENT I. EFFECT OF MASKING NOISE ON FUSION OF A TRAIN OF CLICKS WITH A TRAIN OF NOISE BURSTS}

4.2.1 \textbf{EXPERIMENTAL PROCEDURE}

The stimulus consisted of a train of clicks into the left earphone and a train of noise bursts into the right earphone. (Figure 4a). The independent variable was the level of continuous wide-band masking noise added to one or both sides. When masking noise was added to both sides it was at the same sensation level at both sides. All noise sources were uncorrelated.
FIGURE 4

Schematic representation of stimuli used in experiments.

$1/T$ = Repetition Rate

$\tau_B$ = Noise Burst Duration

$\tau_{RL}$ = Interaural Time Delay (Right Ear Leading)
The noise burst is thermal noise from a General Radio random-noise generator, 20-kc spectrum, gated through a Western Electric 291A mercury relay. The contact arrangement on the relay and the multivibrator used to operate the relay are shown in Figure 5. This contact arrangement eliminated any audible switching transients.

When the experimenter turned on the click and noise burst, the subject adjusted the interval between the click and the onset of the noise burst to obtain a centered image. When he had obtained a centered image, he signalled the experimenter, who recorded the interval between the click and the noise burst. If the subject could not get a centered image, the experimenter recorded "no fusion". The experimenter then turned off the click and the noise burst, adjusted the masking noise to a different intensity, and turned the click and the noise burst back on. We kept the masking noise on between trials. The subject used a buzzer to communicate with the experimenter.

A day's run consisted of five trials (four for subject JP) at each masking noise level being considered. The masking noise level and the side masked were changed at random, to minimize fatigue effects. A run took between a half an hour and an hour and a half to complete. If the run took more than about three quarters
FIGURE 5

Noise burst gate.

(a) Multivibrator used to drive relay.
(b) Contact arrangement of relay.
of an hour, the subject was given a ten-minute break. Each run was repeated on three different days.

The burst lengths $\tau_B$ and repetition rates $1/T$ we investigated were 50-msec, 20-msec, and 10-msec noise bursts at 5 pps, and 10-msec noise bursts at 20 pps. We determined thresholds at 5 pps, using a 50-msec noise burst. The intensity at which a click or a noise burst is just audible depends on the noise burst length and on the repetition rate, but we did not compensate for this when we changed the stimuli. We kept the intensity of the stimuli unchanged for all conditions, to provide a stable reference level.

4.2.2 RESULTS

The results of the experiment are summarized in Figures 6 and 7. Each point is based on fifteen trials (twelve for subject JP). Figure 6 shows the interval between the click and the noise burst for a centered image. Figure 7 shows the proportion of trials in which the subject was able to get a centered image.

With masking noise only in the ear receiving noise bursts, the level of masking noise necessary to prevent fusion was on the average 7 db less than the level of masking noise necessary to make the noise burst inaudible. With masking noise only in the ear
Experiment I. Train of clicks in left ear, train of noise bursts in right ear. Effect of masking noise on $\tau_{RL}$. 

FIGURE 6 (1)
FIGURE 6 (11)
FIGURE 7

Experiment I. Train of clicks in left ear, train of noise bursts in right ear. Effect of masking noise in preventing fusion.
receiving clicks, the level of masking noise necessary to prevent fusion was on the average 1 db less than the level necessary to make the clicks inaudible. With masking noise in both ears, the level of masking noise necessary to prevent fusion was on the average 1.3 db greater than with masking noise only in the ear receiving bursts of noise.

In every case, with masking noise at 0 db SL in both ears, the subject put the noise burst from 2 to 4 msec ahead of the click for a centered image. As masking noise was added to the noise burst, this gap increased. When masking noise was added to the click, it was necessary to present the click earlier relative to the noise burst to obtain centering.

Table I lists the standard deviation of delay between click and noise burst for a centered image. (The average over all levels of masking noise considered is given. In general, the standard deviation increased as masking noise approached the level that would prevent fusion.) Table I also lists the level of masking noise necessary to prevent fusion in 50 per cent of the trials, and the level of masking noise necessary to make the stimulus inaudible monaurally.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Stimulus (Burst length, msec/pps)</th>
<th>Side presented with masking noise</th>
<th>Standard deviation of delay for centered image</th>
<th>Level of masking noise to prevent fusion in 50% of trials</th>
<th>Level of masking noise to make stimulus inaudible</th>
</tr>
</thead>
<tbody>
<tr>
<td>GH</td>
<td>50/5</td>
<td>R</td>
<td>1.9 msec</td>
<td>17.5 dB SL</td>
<td>26 dB SL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10/5</td>
<td>R</td>
<td>1.2</td>
<td>18</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>0.8</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>1.0</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>10/5</td>
<td>R</td>
<td>1.6</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>1.8</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>1.5</td>
<td></td>
<td>31</td>
</tr>
<tr>
<td>JH</td>
<td>50/5</td>
<td>R</td>
<td>1.4</td>
<td>13.5</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>0.9</td>
<td>30.5</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10/5</td>
<td>R</td>
<td>0.9</td>
<td>17</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>0.7</td>
<td>31.5</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10/20</td>
<td>R</td>
<td>1.2</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>0.6</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JP</td>
<td>20/5</td>
<td>R</td>
<td>0.7</td>
<td>19</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>0.5</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10/20</td>
<td>R</td>
<td>1.2</td>
<td>20.5</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>0.6</td>
<td>39.5</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE I**

Experiment I. Train of clicks in left ear, train of noise bursts in right ear.
4.3 EXPERIMENT II. EFFECT OF MASKING NOISE ON FUSION OF TWO TRAINS OF NOISE BURSTS

4.3.1 EXPERIMENTAL PROCEDURE

The stimulus consisted of trains of 20-msec noise bursts into each earphone (Figure 4b). The independent variable was the level of continuous wide-band masking noise added to one or both sides. All four noise sources (noise bursts at the left and the right and masking noise at the left and the right) were uncorrelated. The noise bursts were presented at a level of 20 db SL, at a rate of 5 per second. Each noise burst was generated by gating wide-band random noise from a General Radio random-noise generator through a Western Electric 291A mercury relay, as in Experiment I. Two Scott random-noise generators were used to provide the masking noise.

The subject was required to adjust the relative time of arrival of the two noise bursts to get a centered image. The experimenter recorded whether or not fusion was possible. If fusion was possible, the experimenter recorded the interval between the two noise bursts for a centered image. Five readings were taken at each masking level, with masking noise in the left ear, with masking noise varied from a level at which fusion was never possible to a level at which it was always possible. Corresponding data were then taken with masking
noise in the right ear. Finally, we added masking noise to both ears. In this last case, the difference in the masking levels in the left and right ear was kept equal to the difference between the levels that would prevent fusion monaurally.

4.3.2 RESULTS

Figure 8 shows the proportion of trials in which the subject was able to get a centered image. The levels of masking noise necessary to prevent fusion and necessary to make the noise burst inaudible are tabulated in Table II.

With masking noise in only one ear, the level of masking noise necessary to prevent fusion was on the average 2 db below the level of masking noise that would make the noise burst inaudible monaurally. With masking noise in both ears, the level of masking noise necessary to prevent fusion increased in some trials and decreased in others. We do not consider the change to be significant.

4.4 EXPERIMENT III. EFFECT OF MASKING NOISE ON FUSION OF A TRAIN OF SINGLE CLICKS WITH A TRAIN OF CLOSELY SPACED PAIRS OF CLICKS

4.4.1 EXPERIMENTAL PROCEDURE

The stimulus consisted of a repetitive train of single clicks (click A) into the right ear and a train of two closely spaced clicks (clicks B and C)
Experiment II. Train of noise bursts in left ear, train of noise bursts in right ear. Effect of masking noise in preventing fusion.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Trial number</th>
<th>Side presented with masking noise</th>
<th>Level of masking noise to prevent fusion in 50% of trials</th>
<th>Level of masking noise to make stimulus inaudible</th>
</tr>
</thead>
<tbody>
<tr>
<td>GH</td>
<td>1.</td>
<td>L</td>
<td>25 db SL</td>
<td>25 db SL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>21</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>2.</td>
<td>L</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>3.</td>
<td>L</td>
<td>28</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td>JH</td>
<td>1.</td>
<td>L</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.</td>
<td>L</td>
<td>27</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>3.</td>
<td>L</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>JP</td>
<td>1.</td>
<td>L</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>2.</td>
<td>L</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>30</td>
<td>32</td>
</tr>
</tbody>
</table>

**TABLE II**

Experiment II. Train of noise bursts in left ear, train of noise bursts in right ear.
into the left ear (Figure 4c). The independent variable was the level of masking noise added to the left side. In each trial, the subject attempted to get a centered image by varying the relative timing of click A. If \( \tau_{BC} \) (the interval between click B and click C) was sufficiently large, there were two relative positions of click A which gave centering. If \( \tau_{BC} \) was too small, only one relative position of click A gave centering.

The experimenter varied \( \tau_{BC} \) in .5-msec steps from an interval for which the subject was consistently able to get centering with two positions of click A to an interval for which the subject was able to get centering with two positions of click A in less than 25% of the trials.

The clicks were held at a level of 20 db SL throughout. The experiment was performed for subject JP at a repetition rate of 8 pps, with masking noise at -40 db SL, 15 db SL, and 20 db SL, and for subject JH at 20 pps and 8 pps, with masking noise at -40 db SL and 20 db SL.

4.4.2 RESULTS

Figure 9 shows the proportion of trials in which the subject was able to get two centering positions of click A, for the various conditions. The results indicate that the addition of masking noise reduces the B-C interval at which the subject is able
FIGURE 9

Experiment III. Train of single clicks in left ear, train of pairs of clicks in right ear. Proportion of trials in which subject is able to get two centering positions of single click.
to get two centering positions of click A. Table III lists the B-C intervals at which the subject was able to get two centering positions of click A in 50% of the trials.

<table>
<thead>
<tr>
<th>Repetition Rate</th>
<th>-40 db SL</th>
<th>15 db SL</th>
<th>20 db SL</th>
</tr>
</thead>
<tbody>
<tr>
<td>JH 8 pps</td>
<td>3.9 msec</td>
<td>--</td>
<td>3.75 msec</td>
</tr>
<tr>
<td>20 pps</td>
<td>3.45 msec</td>
<td>--</td>
<td>2.75 msec</td>
</tr>
<tr>
<td>JP 8 pps</td>
<td>5.35 msec</td>
<td>5.25 msec</td>
<td>4.8 msec</td>
</tr>
</tbody>
</table>

TABLE III

Experiment III: Train of single clicks in right ear, train of pairs of clicks in left ear. Interval at which subject is able to get two centering positions of single click in 50% of trials.

4.5 EXPERIMENT IV. EFFECT OF NOISE BURST DURATION ON FUSION OF A TRAIN OF NOISE BURSTS WITH A TRAIN OF CLICKS

4.5.1 EXPERIMENTAL PROCEDURE

The stimulus consisted of a train of clicks in the left ear, and a train of noise bursts in the right ear (Figure 4d). The noise burst duration $\tau_B$ and the repetition rate $1/T$ were the independent variables. The intensity of the noise burst was adjusted so that when the noise burst duration was 1 msec and the repetition rate was 20 pps, the subject made the click and the noise burst arrive at the same time for a centered image.
The click was a 100-\(\mu\)sec rectangular pulse produced by the click generator. The noise burst was produced by gating random noise from a General Radio random-noise generator through a transistorized electronic switch. This switch is shown in Figure 10. It consists of a monostable multivibrator triggered by the sync pulse from the click generator. In the "off" state, the output transistor is in saturation. The impedance from collector to ground is low, and the voltage divider consisting of resistor \(R\) and the output transistor shorts out the noise input. In the "on" state, the output transistor is driven to cutoff. The impedance from collector to ground is high, and the noise input appears at the collector. The biasing circuit on the output transistor makes it possible to adjust the "off" collector voltage so as to eliminate any switching transients. The signal-to-noise ratio of this switch is about 30 db.

The procedure for gathering data in this experiment was the same as that used in Experiments I and II. When the experimenter turned on the click and the noise burst, the subject adjusted the interval between them to get a centered image. If fusion was possible, the experimenter recorded the interval between the click and the noise burst for a centered image.
FIGURE 10

Electronic noise burst gate
We took data at repetition rates of 10 pps, 20 pps, and 50 pps. In a single run, we investigated only one repetition rate. Burst lengths in each run were varied at random. We took five readings for each condition. Only one subject participated in the experiment.

4.5.2 RESULTS

Figure 11 shows the average interval between the noise burst and the click for a centered image ($\tau_{RL}$; positive when the noise burst precedes the click) as a function of noise burst duration $\tau_B$. The spread of the data was high. The five readings taken for any one condition were distributed over a 0.2- to 0.4-msec range. This spread was highest for short noise bursts. However, a definite trend is evident. As burst duration increases, it is necessary to present the noise burst later relative to the click to obtain a centering ($\Delta \tau_{RL}/\Delta \tau_B$ is negative). Beyond a certain burst duration, which depends on the repetition rate, this effect reverses, and as burst duration increases it is necessary to present the noise burst earlier relative to the click to obtain centering ($\Delta \tau_{RL}/\Delta \tau_B$ is positive).

The burst duration at which $\Delta \tau_{RL}/\Delta \tau_B$ changes from negative to positive depends on the repetition rate. The lower the repetition rate (i.e., the
FIGURE 11

Experiment IV. Train of clicks in left ear, train of noise bursts in right ear. Effect of burst length on interval between click and noise burst for a centered image.
longer the period), the longer is this burst duration. This burst duration is about .5 to 1 msec at 50 pps, 2 msec at 20 pps, and 5 msec at 10 pps.

The interval between the noise burst and the click appears to be independent of repetition rate when the noise bursts are short (of the order of 1 msec). Possibly $\tau_{RL}$ is greater at 50 pps than it is at 20 pps or 10 pps, but the spread of the data for very short noise bursts is so great that this is not definite. When the noise bursts are long (of the order of 10 msec), $\tau_{RL}$ increases with increasing repetition rate.
V. INTERPRETATION AND DISCUSSION

5.1 PROPOSED MODEL

On the basis of the above experiments, we propose the following model of the mechanism of binaural fusion: The acoustic stimulus arriving at the ear evokes a timing signal. The strength of the timing signal is determined by the short-term integrated neural response to the stimulus. The timing signals from the two sides travel to a centrally located fusion mechanism. This fusion mechanism compares the signals from the two sides, and under the proper conditions produces the subjective impression of a fused virtual image.

Necessary conditions for fusion are: 1) The discrepancy in relative time of arrival and relative intensity of the timing signals from the two sides must not be too great. 2) The timing signals from both sides must be stronger than some absolute minimum.

In this model, the timing signal is specified by just one quantity, referred to as the "strength" of the timing signal. While the results of the above experiments can be accounted for by this simple model, difficulties are encountered when we try to extend the model to some other binaural phenomena. Fusion of high-frequency pure tones cannot be accounted for satisfactorily by this model. Also, this model does not consider the apparent impossibility of fusion of a high-pass noise burst with a low-pass noise burst.
5.2 INTERPRETATION OF EXPERIMENTAL RESULTS IN TERMS OF PROPOSED MODEL

We now proceed to analyze the experimental results in terms of the above model.

5.2.1 EXPERIMENT I

Supporting evidence for this model comes from two aspects of Experiment I: 1) Although the click and the noise burst were both at 20 db SL, the subject always made the noise burst arrive before the click for a centered image in the absence of masking noise. 2) With masking noise in the left (click) ear, if enough masking noise was added to make fusion impossible, the click was also made inaudible. With masking noise in the right (noise burst) ear, more masking noise was required to make the noise burst inaudible than to make fusion impossible.

1) The apparent trading relationship between time and intensity for fusion gives us a qualitative measure of the relative strength of two signals for fusion. We present the two signals at the two ears of a listener, and have him adjust the relative time of arrival of the two signals to obtain a centered image. In terms of our model, if the listener has signal A precede signal B for a centered image, then we can say that signal B evokes a stronger timing signal than signal A.
When the stimulus consisted of a click at 20 db SL in one ear and a noise burst at 20 db SL in the other, with masking noise at 0 db SL in both ears, the subject put the noise burst at least 2 msec ahead of the click for a centered image. From this we can say that the click evokes a stronger timing signal than the noise burst.

The fact that the click evokes a stronger timing signal than the noise burst, even though the click and the noise burst are at the same sensation level, can be accounted for as follows: The sound-pressure level at which a stimulus is just audible depends upon the stimulus duration. Garner\textsuperscript{14} found that for noise bursts shorter than 100 msec this relation could be expressed by the equation $I \times t^{.8} = C$, where $I$ is the noise intensity at threshold, $t$ is the duration of the noise burst, and $C$ is a constant. Because of this, a 100-\textmu sec click 20 db above threshold will be more intense than a 50-\textmu sec noise burst 20 db above threshold. If the strength of the timing signal is not affected by duration of the stimulus, then the click will evoke a stronger timing signal than the less intense noise burst.

If the strength of the timing signal is not affected by the duration of the stimulus, then the strength of the timing signal must be completely
determined in a very short interval. Referring this to the Goldstein-Kiang model of neural activity, it appears that the strength of the timing signal is determined by the short-term integrated neural response.

2) Consideration of the levels of masking noise necessary to prevent fusion leads to this same conclusion. According to the Goldstein-Kiang model,\(^{18}\) addition of masking noise to a noise burst desynchronizes the neural response, while addition of masking noise to a click reduces the number of neural units firing without desynchronizing the response.

When we added enough masking noise to the noise burst to prevent fusion, we found that the noise burst was still audible. This is the result that would be expected if the strength of the timing signal were determined by the total number of neural units firing over a short interval, and the audibility of the signal were determined by the total number of neural units firing over a much longer interval. If the addition of masking noise to the noise burst spreads the response out over an interval long compared to the integration time of the timing signal, but short compared to the integration time of the mechanism determining whether or not the signal is audible, we would expect that there would be a range of masking noise level for which fusion would be impossible even though the signal was audible.
When we added enough masking noise to the click to prevent fusion, we also made the click inaudible. This also is the result we would predict on the basis of the above assumptions. The addition of masking noise to a click would reduce the number of units firing, but the firing times would still be synchronous. Therefore, the short-term integrated response would be just as strong as the long-term integrated response, and the same level of masking noise that would prevent fusion would make the stimulus inaudible.

This model differs from the model proposed by Deatherage and Hirsch\textsuperscript{10} in that the presence of an $N_1$ response at the round window is not a necessary condition for fusion. The strength of the timing signal in the present model is determined by the short-term integrated response at the round window. While the presence of an $N_1$ response indicates that there is a highly synchronous response, and so is a sufficient condition for fusion, it is not required that the response be so highly synchronous as to produce an $N_1$ response.

While the results of Experiment I are consistent with the prediction of our model, they can also be accounted for in another way. According to the model, a necessary condition for fusion is that the discrepancy between the strengths of the timing signals from the two sides be small. The fact that fusion is impossible
does not necessarily indicate that the timing signal is completely absent. If the unmasked click evokes a stronger timing signal than the unmasked noise burst, as seems to be the case, it is possible that the addition of masking noise to the noise burst would prevent fusion by weakening, but not eradicating, the timing signal. Therefore, there might be a level of masking noise that would prevent fusion without making the noise burst inaudible even if the integration time of the timing signal were the same as the integration time of the mechanism determining whether or not the signal was audible.

5.2.2 EXPERIMENT II

The results of this experiment indicate that the following is a necessary condition for fusion: The discrepancy in relative strength of the timing signals from the two sides must not be too great.

In Experiment I, we found that when a noise burst is fused with a click at the same sensation level, the amount of masking noise (at the ear with the noise bursts) necessary to prevent fusion was on the average 7 db less than the amount of masking noise necessary to make the noise burst inaudible. In the present experiment, we find that when a noise burst is fused with another noise burst at the same sensation level, the amount of masking noise necessary to prevent fusion is on the average only 2 db less than the amount of
masking noise necessary to make the noise burst inaudible.

This can be explained as follows: Fusion becomes impossible when the difference between the strengths of the timing signals from the two sides exceeds a maximum limit. Since the unmasked click evokes a stronger timing signal than the unmasked noise burst, a stronger timing signal is necessary from the masked noise burst when the stimulus in the opposite ear is a click than when it is a noise burst.

5.2.3 EXPERIMENT III

The reduction of the minimum perceptible interval between two clicks by the addition of masking noise was exactly the result we predicted on the basis of the Goldstein-Kiang model of neural activity. The interpretation is this: Neural units are refractory for some time after they fire. That is, they exhibit a raised threshold. In the absence of masking noise, many units are fired by click B and are refractory for some time after. If click C follows too closely after click B, the response to click B is much stronger than the response to click C. When masking noise is added to the pair of clicks, however, it continually renders neural units refractory. Some units refractory at the onset of click B recover enough to fire by the onset of click C. The effect of masking noise is to even out the responses to the two clicks, and in this way to reduce the minimum perceptible interval.
David has reported a closely analogous phenomenon. (It was because of this work of David's that the above experiment was carried out.) The stimulus is identical to that which we used in the above experiment: A train of single pulses in one ear and a train of pairs of pulses in the other ear. Instead of adding masking noise to the pairs of pulses, David increased the repetition rate of both trains, and found that this resulted in a reduction of the minimum perceptible interval. This can be explained in terms of Goldstein's model as follows: At low repetition rates, there is a large interval between click C of one group and click B of the following group. Neural units have a long time to recover, so many neural units fire at the advent of click B. This makes many neural units refractory at the advent of the next C click. At high repetition rates, the interval between click C of one group and click B of the following group is small, so fewer units fire at the advent of click B. Some units that are refractory at the advent of click B recover enough by the advent of click C to fire. At high repetition rates, the silent interval preceding click B is small, so click B loses its advantage over click C.

It would be interesting to observe the neural response in cats to the above stimuli. On the basis of Goldstein and Kiang's model, and on the basis of the two
experiments described above, we would predict that when the stimulus is a train of pairs of clicks, the response at the cochlea to the first and second of a pair of clicks can be made more nearly of the same size either by adding masking noise or by increasing the repetition rate.

5.2.4 EXPERIMENT IV

According to our model, the strength of the timing signal is determined by the total number of neural units firing over a short interval. We shall call this interval the integration time of the timing signal. The original purpose of Experiment IV was to determine the integration time by fusing a click with a noise burst, and varying the duration of the noise burst. The anticipated result was that for noise burst durations shorter than the integration time of the timing signal, the subject would delay the arrival time of the noise burst relative to the click for a centered image when the burst duration was increased. That is, an increase in burst duration would result in an increase in strength of the timing signal from the noise burst. When the noise burst was longer than the integration time of the timing signal, an increase in burst duration would have no effect on the relative time of arrival of the noise burst and click for a centered image.
The observed experimental results did not contradict the predicted results, but they did show that the prediction was an oversimplification. Two complications arose:

1) For short noise bursts, the subject did delay the arrival time of the noise burst relative to the click as burst duration was increased. This was in agreement with the prediction, but as the burst duration became very short the variability of response became very high. This is attributed to the variation in amplitude of individual noise bursts introduced by taking very short samples. The impression reported by the subject for very short noise bursts was of a scattering of apparent sources, so that it was impossible to locate the apparent source in the center of the head.

2) As the noise burst duration was increased, the delay of arrival time of the noise burst relative to the click did not simply level off, but rather began to increase. This is attributed to reduction of response to one noise burst by the presence of the preceding noise burst. In terms of our model, some units responding to one burst were still refractory at the advent of the succeeding burst.

Because of these two complications, it is impossible to state from the above experiment just what
the integration time of the timing signal is. We attempted to eliminate the complication resulting from interaction of noise bursts by presenting the noise bursts at slower repetition rates; but at rates much slower than five per second, subjects found it impossible to obtain a centered image.

5.3 ALTERNATE MODELS

According to the model described above, the fusion mechanism operated simply by comparing the strength of the timing signals from the two sides. The strength of each timing signal is determined by the short-term integrated neural response. This model has the weakness that each timing signal is regarded as a discrete event. This leads to difficulties when we try to extend the model to account for fusion of high-frequency pure tones. A possible extension of the model to account for this might be to postulate some sort of continuous sampling mechanism, of the sort proposed by Békésy.

Another weakness of the model as it stands is that it makes no provision for effects introduced by modifying the spectrum of the stimulus. David, et al., reported no interaction between a 800-cps low-pass and a 6-kc high-pass pulse. Possibly the model proposed by Jeffress, in which the central fusion mechanism compares corresponding regions on the cochleas at the two sides, should be introduced.
VI. CONCLUSIONS

We have conducted a series of psychoacoustic experiments in binaural fusion using impulsive stimuli: clicks and noise bursts. The experiments performed were:
I. Effect of masking noise on fusion of a train of clicks with a train of noise bursts. II. Effect of masking noise on fusion of two trains of noise bursts. III. Effect of masking noise on fusion of a train of single clicks with a train of closely spaced clicks. IV. Effect of noise burst duration on fusion of a train of noise bursts with a train of clicks.

For each condition, we determined whether or not the subject was able to get a centered virtual image. If the subject was able to get a centered virtual image, we measured the relative time of arrival of the stimuli in the two ears.

The relative time of arrival of the stimuli in the two ears was taken as a measure of the relative strength of the timing signals of two stimuli for fusion, in accordance with the time-intensity trading relationship reported by other investigators. The results were interpreted in terms of a simple mathematical model developed by Goldstein and Kiang for interpreting the relative synchrony of neural units responding to these stimuli.

We propose the following model of the mechanism of binaural fusion: Stimuli arriving at the ear evoke
timing signals. The strength of the timing signal is determined by the short-term integrated neural response. A centrally located fusion mechanism compares the strength of the timing signals from the two sides. The model predicts that the subject will judge the image to be centered under the following conditions: 1) The discrepancy between the strengths of the timing signals from the two sides must not be too great, and 2) each timing signal must be stronger than some absolute minimum.

This model is limited to certain types of impulsive stimuli. It cannot be applied in its present form to fusion of high-frequency sinusoids, or to effects introduced by altering the spectrum of the stimulus.
BIBLIOGRAPHY


