The effect of frequency transition on the detection of temporal gaps within a tonal sequence.

Item Type	thesis
Authors	Fitzgibbons, Peter J.
DOI	10.7275/gshq-m254
Download date	2025-04-30 17:18:50
Link to Item	https://hdl.handle.net/20.500.14394/44982

THE EFFECT OF FREQUENCY TRANSITION ON THE DETECTION OF TEMPORAL GAPS WITHIN A TONAL SEQUENCE

A Thesis Presented
By
Peter J. Fitzgibbons

Submitted to the Graduate School of the
University of Massachusetts in partial
fulfillment of the requirements for the degree of

MASTER OF SCIENCE

April 1973

Major Subject Psychology

THE EFFECT OF FREQUENCY TRANSITION ON THE DETECTION OF TEMPORAL GAPS WITHIN A TONAL SEQUENCE

A Thesis

By

Peter J. Fitzgibbons

Approved as to style and content by:

Alexander Pollatsek

Dr. Alexander Pollatsek (Chairman)

Dr. Richard T. Louttit (Head of Department)

Ian B. Thomas

Dr. Ian B. Thomas (Member)

Dr. William Eichelman (Member)

April 1973

ABSTRACT

Previous research has shown that when a sequence of tones which alternates rapidly between two frequency categories is presented to listeners, the perceptual continuity of the pattern is lost. The tones group by frequency domain to form two subjectively co-occurring auditory streams. The phenomenon is described as displaying a "two channel" quality in that listeners must switch their attention rapidly between streams when attempting to track the sequence. In the present experiments listeners attempted to detect brief temporal interruptions between items of a simple auditory sequence, comprised of two high frequency tones followed by two lower frequency tones. Temporal resolution was found to be most difficult when the interruption occurred between tones widely separated in frequency. Results are interpreted as evidence for a processing time delay when shifts of focal attention occur between "sensory channels" developed along the dimensions of auditory frequency.

ACKNOWLEDGMENTS

I wish to express my deepest graditude and appreciation for the guidance and assistance of Dr. Ian B. Thomas and Dr. Alexander Pollatsek. Both persons devoted much of their time and displayed continued enthusiasm during all phases of this research. I would also like to thank Dr. William Eichelman for his time, interest, and stimulating conversation, especially during the initial stages of experimentation. A special note of appreciation is offered to my wife, Kathryn, for her patience and assistance throughout this project.

This research was supported by Grant No. NS-08306 from the National Institute of Health to Dr. Ian B. Thomas.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

INTRODUCTION

The more complex perceptual phenomena cited in the psychological literature have typically reflected on the organizational processes associated with the visual sense modality. To a lesser extent references have been made to auditory perceptual phenomena analagous in complexity to the visual figure-ground relationship, Mach bands, or contour formations in general. The situation is not completely surprising since most auditory patterns are presented sequentially with the most interesting stimuli, speech sounds, not always subject to simple definition. As a result, the few early examples of auditory perceptual organizations were usually found in the musical literature with reference to tonal stimuli.

The most familiar example of an auditory perceptual organization is the musical melody. It may be considered an auditory structure which is not immediately formed, but rather one which finds its completion with the passage of time. Characteristic of many perceptual phenomena, the tonal melody seemingly displays an integral structure not readily apparent through detailed analysis of its tonal ingredients. The general approach to melodic analysis, therefore, has been to investigate its structural flexibility by varying the essential parameters of tempo and tonal frequency. For example, Chandler (1934) stated that the critical requirement for any melody was the existence of small intervals (both frequency and time) between successive notes, along with the repetition of particular notes. The importance of the small frequency interval for perceptual unity had been assumed for some time by musicians since Ortmann (1926) demonstrated that when successive notes deviated too greatly in frequency, the melody seemed to split into two subjectively co-occurring melodies. The higher pitched notes tended to develop a relationship apart from a similar grouping of the lower pitched tones. This perceptual division of a melody can be experienced as the rich polyphonic quality associated with many of the Baroque era compositions, especially in the works of J. S. Bach for solo melodic instruments (Bukofzer, 1947).

More recent reports on the breakdown of melodic integrity have come from Miller and Heise (1950) in a study for which the frequency separation was controlled between two alternating tones, presented at the rate of ten per second. If the frequency separation was less than 15%, listeners reported hearing a "trill,"[1] the continuous gliding of pitch between the two notes to form a unitary auditory structure. With greater frequency separation, the perception of a continuous pattern essentially changed to one of hearing two unrelated, interrupted tones. Later, Heise

---

[1] The visually perceived movement of a light source alternating between two positions in space (phi-phenomenon) seems the direct analogue to the apparent change in pitch for the auditory trill. Bregman and Achim, 1972.

and Miller (1951) presented different tonal patterns, all notes ascending or descending, or ascending initially and then descending, and controlled the frequency of the final or apical note in the respective sequences. Subjective listener reports revealed that whether or not the variable tone was perceived as grouping with others of the pattern, depended on the extent of its deviation from the regular pitch trend in the sequence. That is, the relative, rather than absolute, sequential frequency deviation was critical to melodic integrity. These findings suggest a predictive aspect of the listeners' tracking system with regard to the perceptual formation of melodic contours.

The various subjective statements cited seem to suggest, in a general manner, that not all sequentially presented tonal stimuli are perceived as being part of the same perceptual structure. What is not obvious from such listener reports is the degree of perceptual segregation associated with "split" tonal melodies, under conditions when the phenomenon occurs. Furthermore, to what extent can the perceptual phenomenon be observed to influence the sequential processing of information for other auditory stimuli? Some recent findings are viewed as reflecting on these questions.

Perception of Rapid Auditory Sequences

Norman (1967) discussed experimental results which clearly revealed the need to look beyond time factors alone when considering the temporal course of perception. When two notes (frequency separation within the trill domain) alternated rapidly in a continuous manner, and a third probe tone (30 msec.) was inserted between the two at one instant, listeners could localize the probe only when its frequency was between those of the alternating notes. When the probe frequency fell outside that range, in either direction, listeners were not able to state positively whether it followed the higher or lower pitched note of the trill. Although the probe tone was perceived in all conditions, Norman suggested that the listeners could sequentially process only those items which fell within a certain attentional bandwidth. Critical frequency bands were suggested as probable magnitude estimates of the attentional range. An analogous difficulty is often reported for listeners attempting to localize an audible "click" within a grammatical message (Ladefoged and Broadbent, 1960; Foder and Bever, 1965). While explanations for the latter finding have typically been psycholinguistic in nature, nonlinguistic factors, such as those postulated by Norman, have not been ruled out. For example, Reber and Anderson (1970) reported similar trends in "click migrations" when a speech message is replaced by a series of noise bursts.

Utilizing a different type of auditory sequence, Warren, Obusek, Farmer and Warren (1969) found that listeners experienced difficulty in perceiving the correct temporal order of four qualitatively different sounds. In one condition they presented continuously repeating sequences of four sounds (sine tone, hiss, vowel, and buzz) and found that, for sound segments of 200 msec. duration, temporal ordering performance was at chance level. Segment duration had to increase to 700 msec. before half of the inexperienced listeners could determine the correct order of presentation. These results were quite surprising since the same listeners displayed no difficulty in reporting the sequential order of four vowel sounds for the 200 msec. condition. Thomas, Hill, Carroll, and Bienvenida (1970) subsequently obtained a more complete psychometric function for the temporal ordering of four speech sounds. The vowels employed (/i,/e/,/a/, and /u/) were considered far apart on both articulatory and perceptual domains. With a presentation paradigm similar to that employed by Warren et al. (sounds temporally contiguous and continuously repeating), practiced listeners achieved near perfect performance for segment durations greater than 125 msec. For durations below 100 msec., temporal ordering performance dropped sharply to chance level, even though listeners were able to identify each of the sounds in the sequence.

The observed performance differences between the speech and non-speech sequences were quite surprising to investigators, but so too were the differences between presentation rates for the speech alone. For example, published data reveals an average vowel duration of 200-300 msec. for normal speaking rates (House, 1961). It has been demonstrated, however, that all of the English vowel sounds can be recognized in isolation for durations as brief as 30 msec. (Gray, 1942). In view of these findings, one might expect little difficulty in perceiving the sequential order of four sounds, each 100 msec. in duration. However, when sounds are presented in a sequential manner, as reported for the speech and non-speech patterns, it becomes difficult to predict task performance by adding, in some manner, various process times observed for isolated stimuli. It is reasonable to expect that a practice effect may largely account for the observed superiority of serial recall for speech sounds. The familiarity parameter alone, however, explains neither the poor performance with faster presentation rates nor the perceptual phenomenon experienced with the rapid auditory sequences.

Another approach toward explaining recall variability across types of auditory patterns would be to disregard the formal dichotomy between speech and non-speech stimuli (or any processing dichotomy which assumes a nervous system "wiring" for speech sounds) and view all stimuli patterns with regard to overall perceptual structure. Within this theoretical framework Thomas and Fitzgibbons (1971) attempted to replicate, with tonal stimuli, the psychometric function previously obtained with vowel sequences (Thomas et al., 1970).

Tonal sequences were constructed with notes selected from either a high (H) or low (L) frequency range, or a combination of both H and L segments. Listeners received the patterns at various presentation rates (2 - 11.8 tones/sec.) and were requested to recall sequence order, in accordance with a practiced method of assigning the greatest numerical value to the highest pitched segment. All subjects were allowed to listen to as many repetitions of each four tone sequence as they wished before reporting temporal order. Results indicated that when all sequence tones were drawn from the same frequency category (e.g., all (L), 587, 659, 523, 740 Hz; or all (H), 1976, 1568, 2093, 1760 Hz), the function relating temporal order performance to presentation rate was the same as previously observed with vowel sequences. Performance was near perfect for presentation rates below 8 tones/sec. (segment duration = 125 msec.), but dropped sharply to chance level with faster rates. A quite different finding was observed, however, for the sequence in which the tones alternated between the (H) and (L) categories (e.g., 2093, 523, 1976, 659 Hz). The break in the psychometric function occurred at a much slower presentation rate (2.8 tones/sec.), with many listeners reporting incorrect sequence order at the slowest rate tested (2 tones/sec.). Performance for that type of sequence was clearly more characteristic of the level displayed for non-speech, non-musical sequences of the Warren et al. (1969) study. The difficulty with the alternating (H) and (L) tones was inferred from listener reports to be one purely perceptual in nature. Characteristic of the melodic "splitting" phenomenon previously discussed, listeners claimed to hear two simple melodies for the (H) and (L) tones respectively, and were unable to determine arrival times between the (H) and (L) tones.

A detailed analysis of reported temporal order for tonal sequences characterized by wide jumps in frequency was conducted by Bregman and Campbell (1971). They referred to the perceptual splitting phenomenon as primary auditory stream segregation (PASS). Their listeners displayed great difficulty in perceiving the correct order of arrival for six-tone sequences consisting of three (H) and three (L) tones, which alternated in a varied manner (all sequences presented at 10 tones/sec.). Analysis of data for incorrect serial recall revealed that listener difficulty could be attributed to confusion of order relationships between the (H) and (L) tones. No difficulty was observed for identifying order relationships within either the (H) or (L) triplet subsets. In discussing the results, the authors referred to an auditory stream as a set of events related in a perceptual manner and segregated from other co-occurring streams. When monitoring such tonal groupings, listeners necessarily had to switch their attention from one stream to another to determine temporal order of arrival.

Parameters Affecting Auditory Perceptual Structure

The results from studies dealing with the perception of rapidly presented auditory patterns certainly suggest that the underlying perceptual structure of any sequence can influence the temporal course of information processing. Whether offered by singular findings, or inferred from collective results, the following parameters appear to influence the development of perceptual structure:

Pattern frequency composition

Stimulus frequency appears to be the parameter underlying the most basic of auditory structures. Sequential auditory patterns, characterized by little variability in segment frequency, were perceived as single structures, or melodies. Listeners were readily able to determine temporal relationships between items for these "within-stream" event sets. With regard to the vowel sequences employed, their within-stream nature may have resulted from the methodological concern to maintain a constant fundamental voicing frequency across all vowel samples (Thomas *et al.*, 1970).

Presentation rate

For most of the studies mentioned above, presentation rate was confounded with segment duration since no interstimulus intervals were employed in the auditory sequence. Bregman and Campbell (1971) reported, however, that presentation rate interacts with frequency separation in an inverse relation to produce multiple perceptual groupings from a single tonal sequence. It was also noted in the Thomas and Fitzgibbons (1971) study that even within-stream (all (L) or (H)) sequences broke up, perceptually, at the faster presentation rates.

Nature and relative extent of frequency transitions

Although not specifically discussed in the studies cited, a slower rate of frequency change between successive items of a sequence would be expected to inhibit the perceptual formation of multiple streams.[2] A common example of this feature is observed in the normal speech message with its characteristicly more gradual formant frequency transitions between phonemic elements. It should also be recalled that for certain tonal patterns (Heise and Miller, 1951), the critical frequency separation for perceptual segmentation was a relative one, depending upon the trend of prior frequency intervals. Undoubtedly, research concerning the nature of auditory contour development must deal with this principle; one which essentially embodies the expectancy versus recency controversy with regard to the speed of sequential information processing.

Listening duration

The role played by the listening duration parameter (i.e., the number of sequence repetitions) is not as obvious, except to state that its effect on task performance is opposite in nature to the common practice effect.

---

[2] Bregman, A.L.; Personal Communication, 1972

Especially evident for conditions when the opposing effects of presentation rate and frequency separation are in "balance," continued listening to tonal patterns resulted in the familiar segregation of parts from the perceptual whole. An example of a "balanced" condition was noticed for presentation rates associated with the break point in the psychometric function for the all (L) or (H) tonal patterns in the Thomas and Fitzgibbons (1971) study. Although continued listening was allowed in that temporal order study, the probability of observing correctly ordered reports appeared to decrease with the number of sequence repetitions. The latter observation suggests that segregated streams may require time to develop. Certainly, these few parameters cannot be regarded as exhausting the list of variables which may be observed to influence auditory perceptual structure. They do provide, however, a sufficient list so that, with further consideration of psychological parameters, it may be possible to gain a clearer insight into the integral perceptual grouping which underlies the stream segregation phenomenon.

Segregated Streams and Perceptual Processes

While mention has been made of certain influential experimental variables, much less is known with regard to processing explanations of the stream segregation phenomenon. For conditions when frequency grouping is evident, a purely subjective analysis of the phenomenon might include reference to a figure-figure relationship between co-occurring melodic streams. With less passive monitoring of the acoustical events, attending to either of the auditory streams results in the more familiar figure-ground relationship; one stream described as "distant" from the other in some sensory input space. The analogy to visually ambiguous figures is not direct, however, since very little effort seems required to complete the reversal, which is limited only by the time necessary to switch "attention" to the other perceptual pattern. As mentioned previously, Norman's (1967) analysis of the phenomenon made reference to the operation of fixed perceptual filters for listener frequency processing. The filter concept is postulated, though, with somewhat different characteristics, in the attentional theories of Broadbent (1958) and Treisman (1964). Other accounts of filter theory are summarized by Neisser (1967) and Moray (1969). From a strict interpretation of filter theory, it could be argued that each tonal stream occupied a separate sensory channel limited in spectral extent by the frequency bandwidth of the perceptual filter. Exact determinations of the filter bandwidths are not critical from a conceptual viewpoint; either the peripheral estimate of "critical bands," or the more centrally postulated "attentional bandwidths" (Swets, 1963), support the theory. The fixed aspect of filter bandwidths would seem critical, however, since perceptual tonal groupings would be predictable and dependent on frequency composition of the physical auditory sequence. Given that all sequence frequencies fell within the estimated filter bandwidth (arrived on the same sensory channel), a unitary perceptual stream would result. Since no switching between channels would be required in that situation, the single perceptual structure should prevail, so long as the capable rate of information processing remained greater than sequence presentation rate.

Unfortunately, the clearly defined predictions from notions of fixed perceptual filters are not supported by experimental findings. Without the theoretical modification that filter bandwidth varies inversely with presentation rate, it becomes difficult to explain why a "within-channel" melody becomes perceptually segmented with an increase in presentation rate. That a single channel phenomenon is evident at the new presentation rate with reduced tonal frequency variability, would certainly suggest that no overload of channel capacity had occurred. It should also be restated that the tonal segregation phenomenon was observed to depend on listening duration and the relative frequency of tonal segments. Neither finding is adequately explained by fixed perceptual filters without destroying the implied nature of sensory channels.

An alternative account of the segregation phenomenon, which conceptually deals with the developmental aspects of perceptual structures, is favored by Heise and Miller (1951) and Bregman and Campbell (1971). The explanation asserts that tonal streams develop pre-attentively prior to selective processing. The determination of structural development is postulated to be guided by the Gestalt principles of organization (i.e., good form, symmetry, closure, etc.). The argument for a pre-attentive stage of perceptual analysis was offered by Neisser (1967). With this line of reasoning, Bregman and Campbell suggested that auditory streams were organizational entities not subject to simple definition by any single physical property. An important aspect of this alternative explanation is the general release of the stream segregation phenomenon from any unique attachment to the auditory sense modality, as implied by "critical frequency band" accounts. Unfortunately, little is really known, within a quantitative framework, concerning the processing parameters which interact with sensory information to govern the formation of perceptual entities.

Segregated Streams and Sensory Channels

Beyond consideration of the manner by which auditory streams are perceptually constructed, there appears to be little controversy concerning their "two-channel" nature, once developed. Listeners attempting to track the order of arrival of tonal stimuli find the greatest source of difficulty in performing the mental "switch" between stimuli associated with different perceptual streams. It would seem appropriate, then, to investigate the relationship between auditory streams and sensory channels, as referred to in the vernacular of information processing theory. Of particular interest to the present discussion is the assertion that a switch of attention from one sensory channel to another is a time-consuming process.

The major theoretical accounts of sensory channel processing, since Broadbent's (1958) formalization of the concept, have addressed the topic of "selective attention." The research goal was to determine the extent to which an individual could share his processing capacity between two or more inputs arriving over different sensory channels. Less frequently, studies have been conducted to determine the time delay which may occur when an individual must alter the manner in which he analyzes incoming information. Such a time delay may be effected by a variety of hypothesized processes: the change of perceptual filter location in a sensory input space (Broadbent, 1958); a delay necessitated by some alteration of the parameters of facilitation and inhibition (Treisman, 1964); or, more central still, the time needed to effectively alter the "importance weightings" on completely processed information (Deutch and Deutch, 1963).

The difficulty in measuring the time to switch attention can be attributed to at least two theoretical obstacles.

First, a problem exists concerning what defines a perceptual channel. Secondly, an absence of evidence is noted on either a "minimum dwell time," once locked onto a particular channel, or the synchronization between the physical stimulus switch and the individual's mental switch (Moray, 1969). As a result, most experiments were conducted to determine the effects of an attention switch, rather than a measure of the time delay itself. In so doing, the dichotic listening paradigm was employed to test the assertion that the two ears functioned as separate sensory channels.

The early experimental efforts by Broadbent (1954, 1956, 1957a, 1957b) employed the split-span dichotic presentation paradigm in which different auditory stimuli were presented simultaneously to each ear. Findings from these studies suggested that listeners found it difficult to switch between the ears in an attempt to serialize for report the names of digits arriving simultaneously over each input location. Switch time estimates suggested that listeners were not able to switch from one ear channel to another, and back again to the first, during time intervals less than approximately 1.5 seconds. Later, Moray (1960) presented results which cast doubt on the "switching" interpretation of observed listener difficulty in the Broadbent studies. Moray presented spoken digits to both ears, but alternated the time of arrival for each digit between each ear. With no stimuli arriving simultaneously at both ears, listeners were much better able to report the true sequential order of digits, even though they supposedly had to "switch" repeatedly between two sensory channels.

These results prompted Broadbent and Gregory (1961) to suggest that, under some circumstances, the two ears may function as a single perceptual channel. They did maintain, however, that when inputs were alternated between sense modalities (e.g., eye and ear), and listeners are requested to report stimuli in order of arrival, they do so with less accuracy than if allowed to report the auditory stimuli first, and then the visual stimuli. The latter finding served to preserve the theory for selective processing of information which arrived over separate channels.

A more recent analysis of attention switching has revived, conditionally, the original interpretation of the two ears as separate sensory channels (Treisman, 1971). The experiment conducted was essentially similar to Moray's (1960) in which spoken digits were presented to listeners under several conditions(monaurally, binaurally, and alternating dichotic presentations). Although Moray's results had revealed that serial recall of digits, for an alternating dichotic condition, was superior to the split-span condition, Treisman found a significantly lower level of performance for the alternating dichotic condition when compared with the monaural or binaural condition. The difficulty with the alternating inputs also increased with an increase in presentation rate, but not with list length, which suggested that the problem arose at input (perception), rather than in memory. These findings were interpreted as evidence that there is a limit to the rate at which attention can be shifted between the ears. Furthermore, it appeared that such switching between ear channels effectively reduced the time available for perception and storage of stimuli and, therefore, simulates an increase in presentation rate for binaurally presented digits.

If one ignores, for the purpose of discussion, the fact that tonal sequences for the studies cited were binaurally presented at rapid rates, then a formal comparison can be made between the stream segregation phenomenon and the experience associated with dichotic listening. The performance difference for serial recall between the monaural (or binaural) and alternating dichotic listening conditions is similar, respectively, to that observed for "within" and "between" (frequency alternating) stream tonal sequences. This formal analogy, however, tends to overlook the difference between the two phenomena since channel separation associated with segregated auditory streams seems obligatory (not under control of the listener). The latter consideration was pointed out by Bregman and Campbell (1971), and is supported by other dichotic listening results, which demonstrated that the two ears do not always function as separate channels. Day (1968) reported that when the two ears simultaneously received speech stimuli in the same voice, listeners often heard a single fused word. For example, "poduct" in one ear and "roduct" in the other were heard as "product." Treisman (1970) used simultaneously arriving dichotic stimuli with no shared phonemes at all. Even when listeners were asked to report the input to a single ear, their responses involved as many switches between both inputs as had occurred when both stimuli were actually presented to the same ear simultaneously (e.g., a response, "tak" given "taz" and "gik"). The surprising aspect, with regard to the "fusing" of dichotic inputs, is not that listeners are capable of performing an analysis on the fused information, but rather the suggestion (from Treisman's study) that prevention from doing so is not possible. While it may be argued that even the "nonsense" stimuli employed by Treisman required a degree of semantic processing (presumably a more central function), it is also true that speech stimuli were employed in those studies for which results were offered in support of the two ear - two channel notion. At the stage concerned with semantic processing, it is difficult to account for performance differences between binaural and alternating dichotic listening. It must be assumed that the dimension along which the two ears function as independent channels is one of location in space, presumably more dominant for the alternating dichotic presentation, regardless of types of auditory stimuli employed. Fortunately, no similar confusions arise concerning the parameter basis of channel development with the tonal stream segregation phenomenon. The "dichotic" quality of tonal groupings results from a partitioning along the frequency dimension, and listeners demonstrate no ability to fuse the sensory data across channels.

The apparent similarity between segregated tonal streams and Broadbent's conceptual version of sensory channels suggests that tonal sequences are suitable message sets for the study of focal attention switching. Confirmation of such a switch time would support the assertion that subjectively segregated tonal groupings represent independent organizational entities. Furthermore, such evidence would permit subsequent investigations to concentrate on those physical parameters (stimulus duration, frequency transition, etc.) which may affect auditory contour separation, and presumably switch time as well. The experiments which follow were conducted in an attempt to seek evidence supporting the assumption that a focal switching mechanism is operative during the sequential processing of tonal stimuli.

EXPERIMENT I

A preliminary study was conducted to test the hypothesis that a wide discrete frequency discontinuity between temporally contiguous tones, within an auditory sequence, would also be perceived as a temporal discontinuity in the pattern. It was thought that the perception of a temporal break in the physically continuous pattern might result from the necessary time delay effected by the shift of focal attention between divergent frequency groups. To avoid the necessity for listeners to divide their attention between co-occurring frequency streams (segregation phenomenon), the sequence employed consisted of two high frequency tones (H_1, 2093 Hz.; H_2, 2349 Hz.) followed by two lower frequency tones (L_1, 440 Hz.; L_2, 494 Hz.). All tonal segments were 150 msec. in duration (sequence duration = 600 msec.) and presented binaurally through headphones at 80 dB SPL (re: .0002 dynes/cm^2). Initially, all subjects listened to the sequence for which all tones were temporally contiguous, and a one second interval separated repeated presentations of the $H_1H_2L_1L_2$ pattern. No time limit was set for listening duration and subjects were asked to indicate whether or not any tones of the sequence were separated in time, as though there were a break in the continuity of the pattern. While no such physical gap existed, all eight listeners insisted otherwise, and chose the middle location (between H_2

and L_1) as the point of temporal discontinuity. Another condition was then presented to observe the perceptual effect of an actual gap (G) inserted within the sequence. It was expected that listeners would have little difficulty in detecting even brief gaps (e.g., 20 msec.) located "within" either of the two frequency groups (H_1GH_2 or L_1GL_2). It was quite surprising to observe that listeners still chose the middle location, even though the gap was never inserted at that point. Listeners' tendency to select the point of greatest frequency transition was quite compelling since physical gaps of 75 msec. (half as great as tonal durations) were not correctly localized by half the listeners. Three possible explanations for the difficulty in localizing the smaller gaps within either tonal grouping are:

(A) With the insertion of a singular brief gap between H_1 and H_2 or L_1 and L_2, listeners could "detect" the gap, but chose to report the location of the longer psychological gap which may have resulted from a processing "dead time" during an attention switch between H_2 and L_1.

(B) Listeners processed the H_1H_2 and L_1L_2 groups as singular perceptual units, much as the syllable is postulated to function for speech processing. Thus, the listeners may have detected the existence of a gap probe, but localized it between perceptual units.

(C) Finally, listeners may not have detected the gap at all, as the result of the operation (pre-attentively) of the processes underlying the development of Gestalten structures (e.g., good form, or auditory closure). The listeners, unfamiliar with the nature of the probe signal, responded only to the more dominant cue associated with the sequence location of greatest frequency discontinuity.

An experiment was conducted in order to resolve which, if any, of the possible explanations was acceptable. Note that the gap was never actually inserted at the location of frequency discontinuity, which, from the attention shifting hypothesis, is the critical sequence location. If the individual tonal groups function as separate sensory channels, then a gap inserted at the shift point should be most difficult to detect. The latter assumption applies, in particular, for gap intervals less than, or not greatly in excess of, processor switch time. Since no similar switch time is postulated between items within a given channel, it follows that gap detection performance should improve for temporal interruptions within a frequency group.

Method

Stimulus tapes

The main auditory sequence employed in the study consisted of four tonal segments, two high (H) frequency tones (H_1, 2093 Hz., and H_2, 2394 Hz.) and two low frequency tones (L_1, 440 Hz., and L_2, 494 Hz.). The tonal sequence constructed was ordered $H_1H_2L_1L_2$ (corresponding musical notation; $C_7D_7A_4B_4$).

Long samples of each frequency were recorded using a high quality signal generator (HP 650 A) and tape recorder (General Radio 1525). Care was taken to insure constancy of frequency (General Radio 1151-A digital counter) and intensity (true RMS meter, HP 3400A). Sequences were constructed by splicing together segments of tape of appropriate length (and thus, duration). All segments had an intensity steady-state duration of 150 msec. and were spliced at a 45 degree angle so that the offset amplitude decay of each tone coincided with the amplitude build-up of the succeeding tone. The duration of tonal overlap was observed on oscilloscope to be 5 msec., thus, total item durations were 160 msec. The more gradual tonal onsets and offsets were employed to avoid perceptual "clicks" associated with instantaneous end points, when preceded or followed by periods of silence. Nine comparison sequences were constructed by inserting each of three gap (G) tape segments (20, 40, and 80 msec.) into one of three possible locations within the sequence; for example, HGHLL, HHGLL, HHLGL. Each comparison sequence was, therefore, greater in overall duration than a standard sequence for which all tones were temporally contiguous. Another nine tapes were similarly constructed which employed the same tonal elements but reversed in sequential order, $L_2L_1H_2H_1$. A control sequence was then constructed which consisted of four tonal segments all within a range of three semitones ($A^{\#}_4$, B_4, $G^{\#}_4$, A_4).

The sequence was denoted $L_3L_4L_1L_2$ with respective frequencies, 466, 494, 415, 440 Hz. The control sequence displayed a similar underlying pattern to the $H_1H_2L_1L_2$ sequence with the two higher frequencies preceding the lower ones. In fact, the direction of frequency shifts in the control tape was identical to that of the $H_1H_2L_1L_2$ sequence. Three tapes were constructed which employed the control sequence, each with a 20 msec. gap inserted in one of three locations.

In total, twenty-one experimental sequences were employed: nine for the $H_1H_2L_1L_2$ pattern (3 gaps x 3 locations), nine $L_2L_1H_2H_1$ (3 gaps x 3 locations), and three for the $L_3L_4L_1L_2$ control pattern (1 gap x 3 locations). Practice tapes were also constructed which corresponded to the experimental patterns, but contained a 200 msec. gap inserted into one of the sequence locations.

Each tape constructed was subsequently made into a physical loop for which a standard sequence appeared on one track of the tape, and one comparison sequence on the other track. The comparison and standard sequences on each loop were the same with the exception that the comparison sequence contained a single gap in one location.

Testing procedure

Six undergraduate students from the Psychology Department at the University of Massachusetts participated as listeners in the experiment. All students were individually tested and paid for their services. Twelve experimental conditions were conducted with each listener, four tape loops in a given session on three consecutive days. Each session lasted approximately one hour. Three listeners were presented with nine $H_1H_2L_1L_2$ and three $L_3L_4L_1L_2$ sequences; the other three listeners were presented with nine $L_2L_1H_2H_1$ and three $L_3L_4L_1L_2$ sequences. Within a given session each listener was presented with three $H_1H_2L_1L_2$ (or $L_2L_1H_2H_1$) and one $L_3L_4L_1L_2$ sequences. Each tape contained a different gap inserted in a different location. In the other two sessions, the remaining gap durations were presented, so that each subject was tested with all nine combinations of gap duration and location, and also with the 20 msec. gap in the three locations of the $L_3L_4L_1L_2$ sequence. Each subject received the twelve conditions in a different order, and the gap duration presentation order was counterbalanced across the subjects and the sessions of each individual. Listening was done binaurally in phase through headphones (Koss PRO-4) at 80 dB SPL (re: .0002 dynes/cm.^{2}). At that intensity level the tones of the sequence were judged by the experimenter to be equal in perceived loudness.

Fifty trials were conducted for each of the twelve conditions for each subject. A single trial consisted of the presentation of either a standard or comparison sequence, twice, with a 1.2 msec. inter-sequence interval. (It was thought that two sequence presentations would allow for better listener concentration on the critical sequence location.)

The subjects were requested to report verbally "yes" or "no" with regard to whether or not the twice listened-to sequence contained the temporal gap in a previously indicated location. Subjects were encouraged to guess when uncertain. Each block of fifty trials employed a single standard and comparison sequence, with a single gap placed in the same location. Before each block of trials, subjects were presented with the appropriate practice tape, with a 200 msec. gap inserted into the same sequence and location as would subsequently be employed throughout the next fifty trials. The practice trials were conducted until the listener became familiar with the task and sequence type, as evidenced by their meeting a criterion of ten consecutive correct responses. After each practice session, the listener was again instructed about which sequence location might contain the temporal gap, and also informed of a fifty percent probability of occurrence for trials containing the gap. Trial by trial feedback was also given to further prevent non-perceptual factors from entering into listener decisions. Twenty-five "gap" trials were randomly distributed within each experimental block.

Results

A score for each subject was calculated to designate the proportion (P) of correct responses for each experimental block of trials. Analysis of sequence presentation order revealed no significant difference in gap detection performance between the HHLL and LLHH listener groups, F(1,4) = 1.01, p>.05. The mean (P) scores and standard deviations are presented as the call values in Tables 1 and 2, respectively. With the exception of values for the 20 (C) msec. gap duration, the listed means in each column refer to the gap frequency environments associated with the locations of the HHLL sequence presentation order. The listed values in the rows designated by the 20 (C) msec. gap duration refer to LLLL control sequence.

First analysis of variance revealed significant main effects of both gap location, F (2,10) = 17.2,p<.001, and gap duration, F (3,15) = 16.35, p<.001, on the detection of temporal discontinuities within the tonal sequence. These effects are plotted in Figures 1 and 2, respectively. Figure 1 is a graph of Table 1 column values averaged across gap durations. Similarly, Figure 2 represents a graph of Table 1 gap duration values averaged across sequence locations. Further analysis of mean contrasts revealed no significant difference in performance between gap locations one and three (Fig. 1), the "within" category frequency groups, F (1,10) = 3.18, p>.05. Performance for gap location two, however, was significantly poorer than the average results for locations one and three, F (1,10) = 32.3, p<.001. Gap location two corresponds to the "between" frequency category position for the HHLL sequence. A similar comparison of values displayed in Figure 2 revealed that the demonstrated low point of the graph (20 msec. gap) represented a significantly lower performance level than was average for the other gap durations, F (1,15) = 36.7, p<.001.

The interaction between gap location and gap duration, suggested from observation of Figures 1 and 2, was also found to be significant, F (6,30) = 3.06, p<.025. Figure 3 displays the interaction by plotting the mean performance for each sequence gap location as a function of gap duration. Observation of Figure 3 reveals the high level of detection performance (greater than 90%) for all temporal gaps inserted in locations one and three of the tonal sequences. By comparison of corresponding values for location two, a significant difference was found to exist between performance with the control sequence (LLLL) and the experimental sequence, HHLL, F (1,55) = 12, p<.005. The results indicate that the poor level of gap detection performance associated with the middle sequence location was solely characteristic of the HHLL tonal pattern. No gap detection difficulty was found to exist for any locations of the control sequence.

Observation of results in Table 1 and Figure 3 reveals that, with the exception of the middle location of the HHLL pattern, the listeners' detection performance was quite good for all durations of the inserted gap. The 20 and 40 msec. gaps, at the frequency shift location of the HHLL tonal pattern, presented more difficulty to listeners than did any of the other experimental conditions. Since these results are generally supportive of the attention switching hypothesis, as previously discussed, a second experiment was conducted to investigate the possible perceptual cues utilized by listeners attempting to detect the shorter temporal gaps, within the frequency shifting location of the HHLL sequence.

Table 1

Mean Values of Proportion Correct
for Conditions of Experiment 1.

Gap Duration in msec.	Gap Location		
	HGHLL	HHGLL	HHLGL
20	.950	.670	.910
40	1.000	.840	.990
80	1.000	.940	.990
20 (C)	.960	.920	.910

Table 2

Standard Deviation Values for Conditions
of Experiment 1.

Gap Duration in msec.	Gap Location		
	HGHLL	HHGLL	HHLGL
20	.042	.052	.082
40	.000	.126	.010
80	.000	.084	.023
20 (C)	.053	.048	.051

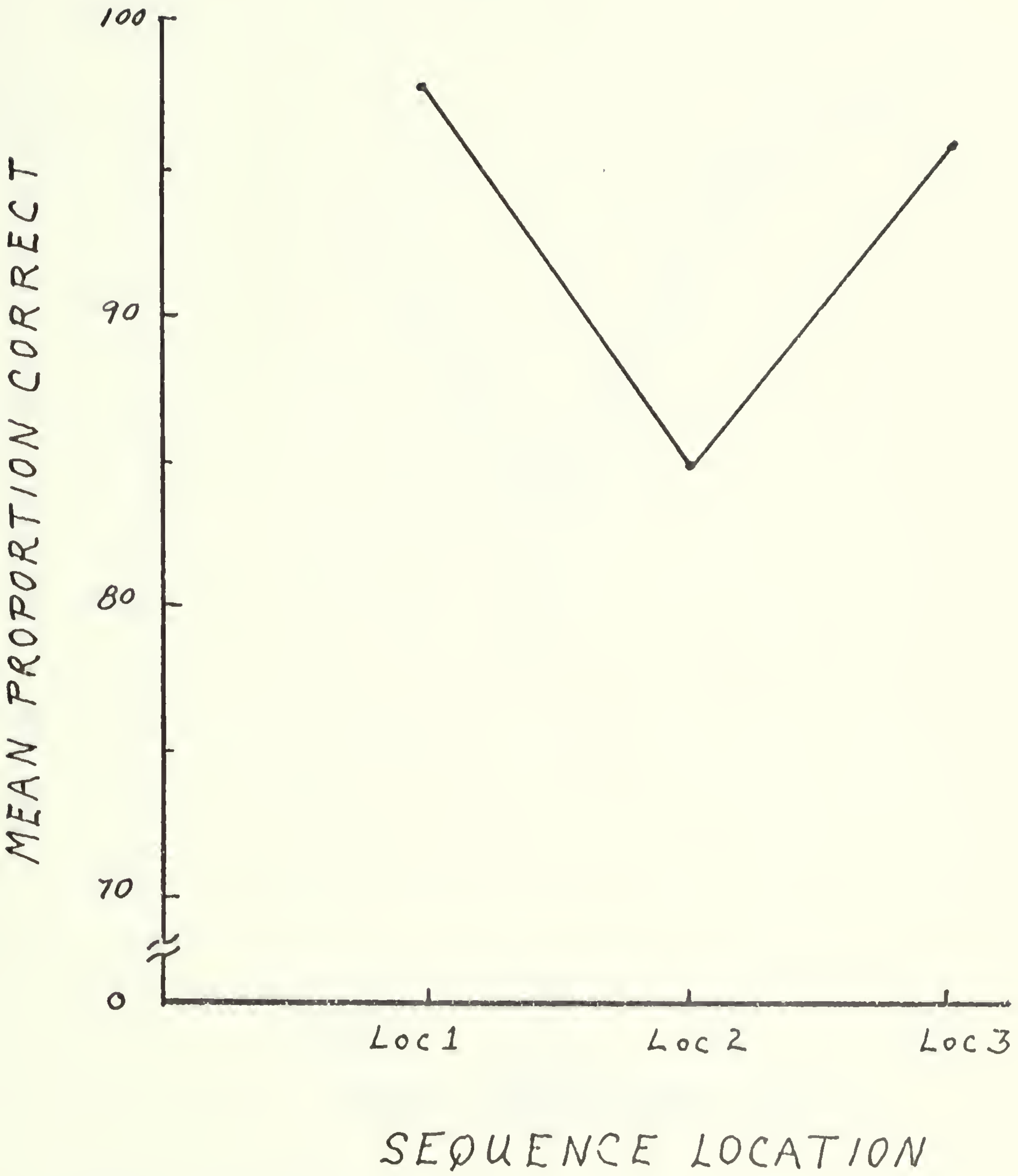

Figure 1. Mean proportion correct as a function of temporal gap location.

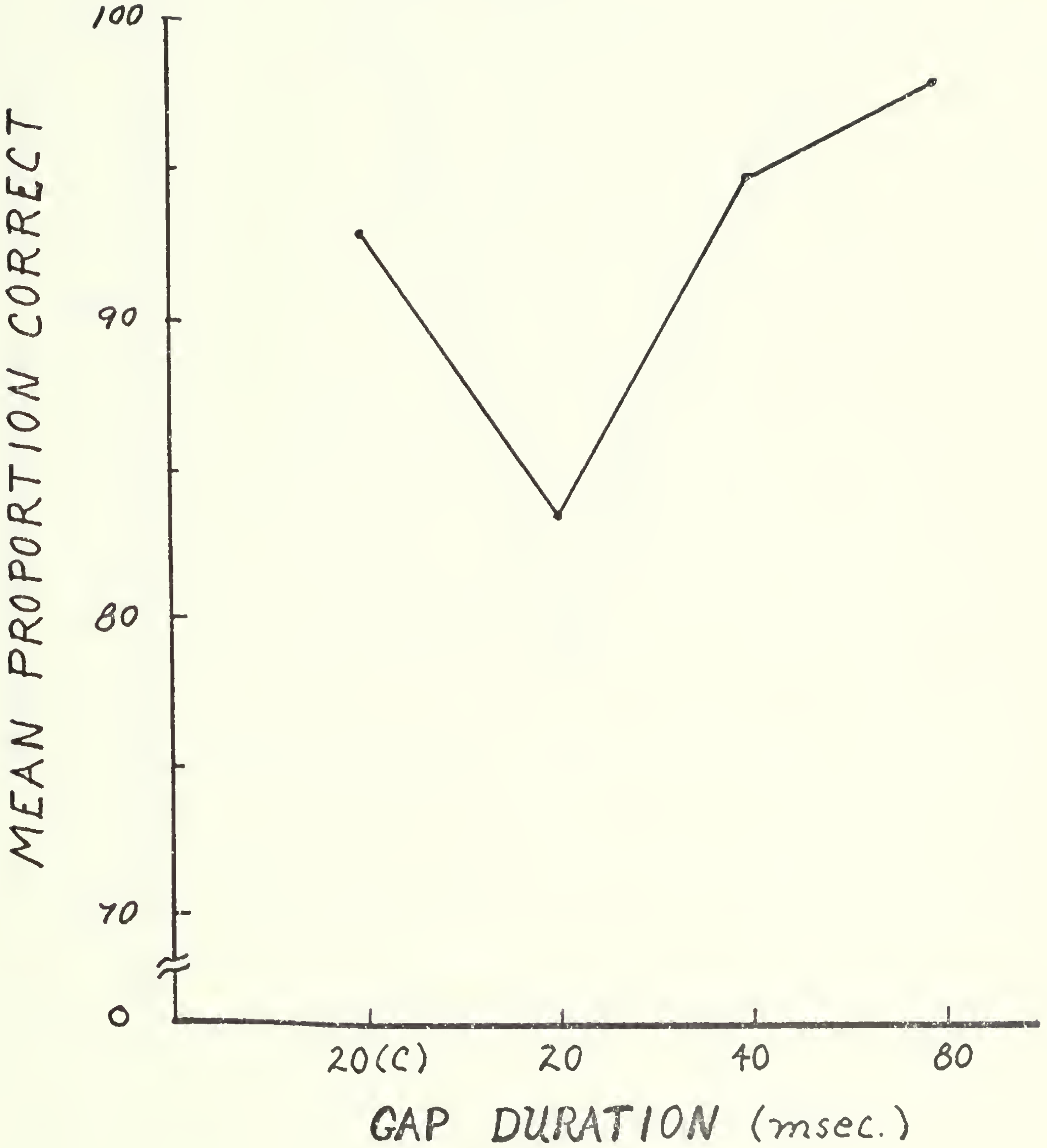

Figure 2. Mean proportion correct as a function of temporal gap duration.

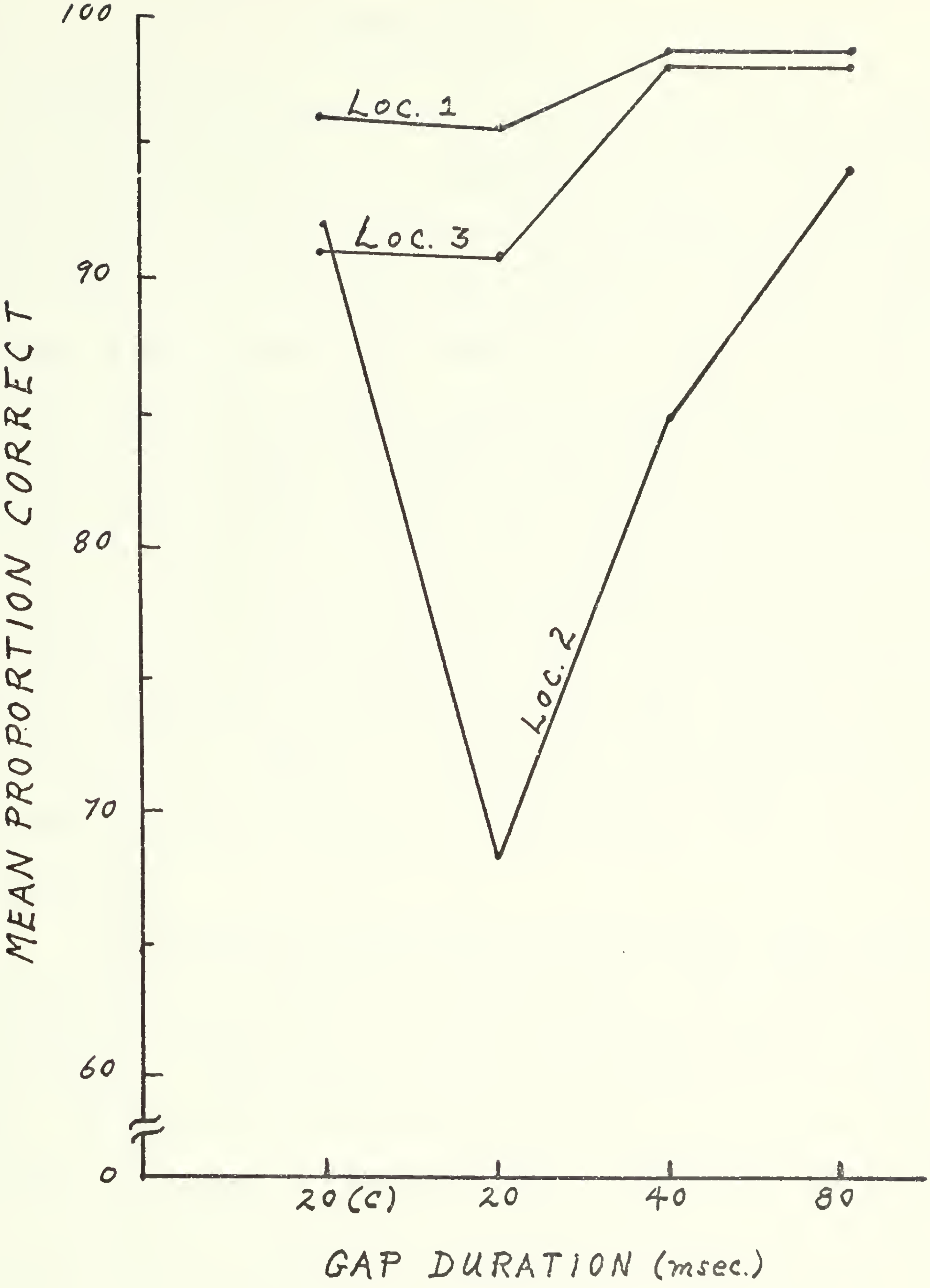

Figure 3. Mean proportion correct as a function of gap duration for each sequence location.

EXPERIMENT 2

Stated quite generally, the required task for listeners in Experiment 1 could be regarded as one of detecting just noticeable changes in tonal rhythm between two auditory patterns, differing only with respect to a brief temporal interruption between two segments. As such, one must conclude that slight interruptions of sequential continuity, at points of abrupt frequency transition, result in less noticeable rhythmic changes than interruptions otherwise located. Since overall performance with the HHGLL (G=20 msec.) condition was somewhat better than chance level (though less than the commonly accepted 75% threshold level), it became of subsequent interest to determine which perceptual cues were employed by listeners to decide in favor of gap existence.

One possible listening strategy states that subjects, attempting to detect the temporal gap, monitored the first frequency group (H_1H_2), and, at the "perceived offset" of H_2, "switched" to process the onset of the lower frequency group (L_1 onset). Within a block of trials, if L_1 onset was detected or perceived as different from its initial onset on other trials, the listener could respond "yes" (for example), and then utilize experimenter feedback to guide response to similar perceptions on subsequent trials. The important feature of this hypothesis (discrimination learning) is that decisions were based on cues obtained during the critical H_2 (offset) - L_1(onset) interval, and not based on cues received from other parts of the tonal sequence. The assumptions involved in the hypothesis would predict that slight duration modifications of either H_2 or L_1 would not affect gap detection performance, since the cues available within the H_2 (offset) - L_1 (onset) interval remain unchanged.

Another explanation of the observed results from Experiment 1 states that the 20 msec. gap was not great enough for subjects to perform the discrimination learning task implied under the first hypothesis. That is, for both the HHLL and HHGLL patterns, the initial perceptions of tone L_1, after the switch, were not discriminably different. Lacking L_1 onset cues, then, another strategy would be to utilize the cue associated with the delayed offset of L_1 in the HHGLL, compared to HHLL, sequence. In this manner listeners could perform a duration discrimination on L_1 <u>after the switch</u>, and compute that the trials of greater perceived L_1 duration must also be the trials on which the gap was present. Comments from some listeners in Experiment 1 suggested the possible utilization of such a strategy. Furthermore, independent evidence on durational discrimination of auditory stimuli reports a differential sensitivity of approximately 10 msec. for durations within the range of present discussion, 100-150 msec. standards. (Small and Campbell, 1961, Abel, 1972). Confirmation of the second hypothesis would support the argument that the better than chance level performance of Experiment 1 listeners resulted from their utilization of information received during the H_2 (offset) - L_1 (offset) interval, rather than during the H_2 (offset) - L_1 (onset) interval. Therefore, a decrease in the duration of tone L_1 would predictably result in the perception of the HHGLL sequence being more similar to that of the unmodified standard HHLL. Small duration modifications of tone H_2, under hypothesis 2 assumptions, would be predicted to have less effect on gap detection, since the H_2 (offset) - L_1 (offset) interval is not altered. The predictions under conditions of H_2 modifications, for both hypotheses, necessarily assume that the <u>time of switch</u> will not be affected -- the time between physical stimulus offset and perceived offset.

Confirmation of hypothesis 2 would lend support to the suggestion by Treisman (1971) that the monitoring of an auditory message, which requires the successive switching of attention, effectively simulates an increase in presentation rate by reducing the time available for perception and storage. Experiment 2 was conducted to test the validity of the above theories of gap detection.

Method

Stimulus tapes

The single tonal sequence $H_1H_2L_1L_2$ served as the standard sequence throughout the experiment. As constructed in Experiment 1, all tones of the sequence were temporally contiguous and 160 msec. in duration. Three experimental tape loops were constructed, each with a different comparison sequence on one track of the tape. Each comparison sequence contained the same tones and presentation order as the standard sequence and contained a 20 msec. gap (G) in the middle sequence location, between H_2 and L_1. Each comparison sequence differed only with respect to the duration of a single tonal element. One condition employed a 20 msec. trunction of tone H_2 before the gap (HH_TGLL), another truncated tone L_1 by 20 msec. ($HHGL_TL$), and the third left unchanged all tonal durations of 160 msec. (HHGLL).

Procedure

Nine listeners participated in the experiment and were paid for their service. None of the subjects had participated in Experiment 1. The same testing procedure of Experiment 1 was employed in this study. Each subject received a block of fifty trials for each of the three experimental conditions. Besides receiving the practice session before each trial block, the listeners were given added practice at detecting 20 and 40 msec. gaps in some of the conditions employed in Experiment 1. Order of condition presentation was counterbalanced across listeners, and each listener received all three conditions in a single one hour session, with a five minute rest period between trial blocks.

Results

Mean proportion correct (P) and standard deviations for each experimental condition are listed in Table 2. Performance between listeners was more varied in this study, F (8,16) = 3.54, $p < .025$, but there was a significant difference in performance across experimental conditions, F (2,16) = 13.6, $p < .001$. While the slight difference between the two "modified" conditions, HH_TGLL and $HHGL_TL$, was not found to be significant, performance for the unmodified condition, HHGLL, was found to represent a significant increase over the average of the other two, F (1,16) = 27.1, $p < .001$. Comparison of the present results to those of Experiment 1 for the unmodified condition (P = .67) reveals the similar difficulties for both listener groups to detect a brief gap between two divergent frequency groups.

The results do not completely support all assumptions of either hypothesis previously mentioned. Hypothesis 1 predicted that listeners would perform equally well for the three experimental conditions. The results from only one listener approximate that predicted outcome. All other listeners, however, performed significantly better on the unmodified sequence HHGLL as compared to the $HHGL_TL$ condition. The latter result is the predicted outcome under the assumption of hypothesis 2 which suggested that listeners were able to utilize the added perceptual duration of tone L_1 (unmodified sequence) to distinguish the standard and comparison sequences. The results on the HH_TGLL condition do not support either hypothesis. In fact, the logical extension of hypothesis 2 would suggest that gap detection for the HH_TGLL condition be the same or better than the HHGLL condition, depending respectively on whether the L_1 duration cue or the L_1 and H_2 duration cues was utilized in the decision process.

Table 3

Mean Values of Proportion Correct
and Standard Deviation Scores for
Conditions of Experiment 2.

Gap Duration in msec.	Sequence Condition		
	HH_TGLL	$HHGL_TL$	HHGLL
20	.570	.560	.695
	(.075)	(.065)	(.087)

(Parenthesized values represent standard deviation scores)

DISCUSSION

The main results from the two studies were quite striking. Temporal resolution of items within a simple tonal grouping was far superior to the resolution of items belonging to different perceptual groups. When all successive tones were part of a unitary perceptual organization, gap detection was observed to be independent of place within the continuum. Findings from Experiment 2 not only replicated those for respective conditions of Experiment 1, but further suggested that when a divergent frequency group follows another without delay, some initial information of the second group does not enter into overall perception of the sequence. In reference to the findings preliminary to conduction of Experiment 1, the first mentioned hypothesis, postulating a processing "silent" interval, effected by a switch of focal attention, is the only explanation which finds support in the observed data. When confusion of possible locations for gap insertion was eliminated, listeners were able to detect the smallest within-stream temporal interruptions with a high degree of accuracy. Performance levels for the HHGL$_T$L condition of Experiment 2 further reveals that initial indicators of between-stream gap detection success (HHGLL, Experiment 1) were inflated in value. That is, listeners in the HHGLL conditions did not detect the gap *per se*, but computed its occurrence from the added information of a perceptually longer tone after the switch, on target trials. For the latter assertion, it is not necessary to assume that all listeners actually performed a duration discrimination on L_1, after the switch, in the standard and comparison sequences. It is only necessary to assume that some perceptual by-product of an L_1 duration increment (e.g., increase in loudness) was utilized to decide in favor of gap occurrence.

An adequate explanation for the similarity of results between the HH_TGLL and $HHGL_TL$ conditions is not available at the present time. It could be argued that both these sequences and their common standard sequence are equivalent with respect to the time interval between H_2(onset) and L_1 (offset), thus accounting for the difficulty in discriminating between them. The argument fails to explain, however, why the same duration increment, which distinguishes HHGLL and LLGLL from their respective standards, is more readily detectable in the latter condition. The same failure, by comparison to within-stream detection performance, results for any other explanations based solely upon listener strategies of discriminating physically measurable time intervals between standard and comparison sequences.

It seems more reasonable to expect that duration changes in H_2, before the temporal gap, affected the <u>time of</u> attention switch by altering the interval after stimulus offset at which listener criterion of <u>perceptual</u> offset was met. To explain present findings (HH_TGLL), it must be assumed that a decrease in tone H_2 duration effectively increased the delay before switching to L_1, thus negating the perceptual effects normally resulting from the inserted time gap. The result of an increased delay before switching would be to enhance the similarity between HH_TGLL and HHLL sequences, especially with regard to tone L_1. This topic is discussed further (below) with regard to the observed relationship between stimulus intensity and the decay of auditory sensation.

Before postulating a processing delay time for transit between divergent frequency contours, one other non-attention shifting explanation of results needs to be considered. A compelling alternative explanation of gap detection results invokes the principle of temporal frequency masking. It might be reasonable to assume that discrete frequency discontinuities are perceived as temporal discontinuities because tone H_2 (in the standard $H_1H_2L_1L_2$) is partially masked in backward direction, resulting in the perception of a temporal gap and shorter perceived duration of H_2. Evidence from studies on temporal frequency masking reveal that lower frequency tones are better maskers of higher frequencies (vice high frequencies masking lows), and backward masking is more powerful than forward masking during the first few msec. (Homick, Elfner and Bothe, 1969; Elliot, 1962). With respect to perception of tone H_2, and the perceived temporal interruption, the similarity between HH_TGLL and HHLL might then be explained by masking. The same explanation, however, cannot account for performance differences across the $\mathrm{HHGL_TL}$ and HHGLL conditions, since Elliot (1964) has demonstrated that the backward masking function is not affected by masker duration.

More serious evidence against a temporal masking explanation comes from the gap detection performance in the $L_3L_4L_1L_2$ condition. In the standard auditory masking experiment, the most effective temporal masking results when tones are close together in frequency. Therefore, if perceived gaps are the by-product of temporal masking, then detection of actual temporal gaps should be more difficult for trial blocks employing the LLLL sequence. Present findings do not support that predicted outcome. Finally, the relative intensity levels employed in the present studies further suggest that masking participation was negligible. Temporal masking functions are usually derived by relating minimum detectable probe intensities to time of occurrence for masking stimuli of much greater intensity. Little perceptual deletion of physical stimulus energy would be expected for probes and maskers of equally great intensity.

The temporal gap detection task was employed, essentially, as a means of obtaining experimental support for the assumption that sensory channels can develop along the frequency dimension for auditory stimuli. With respect to the perceptual experience referred to as tonal stream segregation, the experimental endeavor was one of demonstrating that the phenomenal separation between streams was as "real" as the separation in auditory space for dichotically presented stimuli -- real, that is, with respect to the more central fusion of sensory information. Listeners' inability to distinguish between temporally continuous and discontinuous auditory patterns at points of abrupt frequency transition seems to provide the necessary evidence in support of such assumptions.

It should be clear that no precise estimates of the time taken to switch attention can be gathered from results of the present studies. Furthermore, such estimates would be of doubtful significance in predicting how well listeners would perform in more demanding tasks. For tonal patterns requiring a greater number of tracking switches, one cannot assume a constancy for each attention switching time, or, as previously mentioned, a time synchronization between stimuli offsets and initiations of the mental switches (i.e., an inertia inherent to the focal switching mechanism). The tonal sequence, HHLL, requires a single switch between a pair of simple frequency contours. Even though it could be assumed that listeners were quite familiar with the sequence (they knew "where" to switch), it is difficult to determine when, after H_2 offset, that switches were initiated. A listener attempting to minimize the number of undetected target trials would undoubtedly employ a less stringent criterion for perceptual offset of tone H_2. That is, the listener may initiate his switch on the basis of a minimum decay in sensation level (e.g., j.n.d. for intensity) after H_2 offset. Other listeners, however, may desire more affirmative proof of stimulus offset (greater decay of sensation), but sacrifice vital switch time in the gap interval, thereby reducing their probability of detecting L_1 onset, or its more extensive perception.

In addition to variability of listeners' offset criteria, switch time estimates are further complicated by the variability in the rate of sensation decay effected by stimulus intensity. Since the total time for auditory sensation to decay to threshold has been demonstrated to be constant, and independent of stimulus intensity and frequency (Bekesy 1960; Plomp, 1964), it stands to reason that louder sensations must decay at a faster rate. Any change in the rate of decay would obviously affect perception of stimulus offset, independent of the switching criteria employed by listeners. For these reasons, the observed similarity of results for the HH_TGLL and $HHGL_TL$ experimental conditions is not regarded as significant as the large performance discrepancy between the $HHGL_TL$ and HHGLL sequences. The latter two sequences are identical prior to the inserted gap. It is quite possible, for example, that a decrease in duration of H_2, prior to the gap, affected the terminal sensation level at its offset. Since the growth of loudness with duration is not solely a threshold level phenomenon, the decay of the perceptually less intense H_2 (in

HH_TGLL) would progress at a slower rate than in the other conditions. With a slower decay, and later occurring initiation of the switch to L_1, both standard and comparison sequences had become perceptually less discriminable. The important point, however, is that many stimulus factors need to be investigated which may influence even a single switch of focal attention. It would not be surprising to find that many of these factors are the same as those mentioned as affecting the stream segregation phenomena itself.

Within the last decade, several investigators have employed a temporal gap detection task in attempting to obtain various measures of auditory acuity. Although conducted for differing purposes, with utilization of simple auditory stimuli, the results from such studies can be directly compared to present findings. The remaining discussion is devoted to analysis of certain parameters demonstrated to affect temporal gap detection. In so doing, it is hoped that more significant insight will emerge concerning the developmental aspects of the more complex contours associated with the tonal stream segregation phenomenon.

An interesting finding reported by Elfner and Caskey (1965) can be interpreted as demonstrating the manner by which both stimulus frequency and dichotic listening may be utilized to enhance within-stream processing of temporal gaps. Their investigation concerned analysis of a previously demonstrated phenomenon (Thurlow and Elfner, 1959) in which one of two alternating tones was perceived as continuously present, even though it was interrupted during the periodic presentations of the other tone. Elfner and Caskey presented to listeners a noise signal which was periodically interrupted by a probe tone of variable frequency and duration. "Continuity thresholds" were calculated as the minimum noise interruption interval (probe duration) for which the perception of noise continuity broke down. Monaural and dichotic presentations of noise and probe tone were employed as experimental conditions. Their findings revealed that noise gaps were most easily detected when no probe tone was presented to fill the interrupted intervals (threshold gaps approximately 7 msec.). Gaps were more readily detected for dichotic listening than monaural, but, with an increase in probe frequency, monaural detection performance approached that for the dichotic condition. Both findings are expected from a channel processing viewpoint, particularly if the assumption is made that high frequency contours are more perceptually isolated from a noise stream than are the lower frequencies. The assumption finds physiological support in the observation that lower frequencies, like noise segments, stimulate a greater number of sensory receptors along the basilar membrane than the higher pitched tones. Behavioral evidence also reveals that continuity thresholds for low frequency streams are more similar to those for noise streams than values observed with higher frequencies.

Plomp (1964) first utilized the gap detection paradigm to investigate the decay of auditory sensation (mentioned previously). In a two alternative forced-choice task, his listeners attempted to detect the minimum gap between 200 msec. noise bursts. Threshold detection performance was demonstrated for gaps as brief as 3 msec. when the bordering stimuli were of equal intensity. Consistent with the general finding that auditory sensation decays exponentially in time, gap thresholds increased when the intensity of the preceding noise burst decreased. These results, and those reported by Elfner and Caskey, reveal that temporal gap thresholds for spectrally wideband stimuli are quite small. Although not specifically tested by Plomp, the suggestion was made that similar gap thresholds would be observed for narrowband tonal stimuli.

Recently, Williams and Perrott (1972) employed a gap detection task with pairs of tonal stimuli. Their aim was to obtain a resolution sensitivity measure for comparison with similar indices obtained by other experimental methods. Hirsh (1957), for example, reported that a 20 msec. onset asynchrony between tonal stimuli was necessary for listeners to correctly identify which tone was presented first. The onset delay was found to be independent of the two-tone frequency separation. Measures by the gap detection method, however, revealed quite different findings. Temporal gap thresholds were observed to increase with both tonal duration and frequency separation. The critical effect of duration was observed in the interaction with frequency separation -- the more powerful effect. For example, with shorter tonal durations (3-30 msec.) gap thresholds were nearly constant (2.5-5 msec.) across all frequency separations (0-480 Hz). As tonal durations increased (to 300 msec.), however, the frequency separations for constant gap thresholds became progressively more narrow, and beyond a critical separation, thresholds increased monotonically with tonal frequency divergency. One example revealed a gap threshold of 8 msec. for 300 msec. tones of less than 8 Hz. separation; with a 480 Hz. separation, and no change in duration, the gap threshold increased to 40 msec. Unfortunately, no explanations for these findings were offered by the authors. Certainly, it can be expected that both the intensity (15 dB sensation level) and frequency region (1000 Hz) for stimuli employed had some effect on the absolute thresholds observed. The relative effects of stimuli duration and its interaction with tonal frequency do suggest, however, that perceptual segregation of pitch is a time dependent process.

From a review of the few factors observed to influence listeners' performance in the detection of brief temporal interruptions between auditory stimuli, certain processing generalizations seem evident. Within the context of perceptually segregated auditory patterns, and the assumed relation to attention switching, findings indicate that successively arriving wideband stimuli are processed as single channel events. It seems most probable that brief gap detection thresholds for "noise" stimuli reflect a listener strategy of detecting an onset of nervous activity, subsequent to a perceptual decay in sensation, resulting from offset of stimulation preceeding the temporal gap. Likewise, the similarly brief gap thresholds associated with short duration tonal stimuli suggest negligible participation of an attention shifting factor on detection of temporal interruptions. These "noise-like" gap thresholds observed for short duration tonal stimuli make sense only if one assumes that frequency channels develop as a function of time. Initially, after tone pulse onset, nervous excitation patterns are widespread along the frequency dimension -- independent of the steady-state frequencies presented. This need not imply a "white noise" initial representation for tonal stimuli, since most likely there occurs some initial sorting by bandlimited frequency domains. As tonal (or narrowband) stimulation persists, however, some level of pre-attentive (or at least non-attentive) processing could function to produce finer frequency contours within a pre-selective input space. As such contours developed, for stimuli which are in fact narrowband, it is quite possible that two initially "within-channel" events could become "between-channel" stimuli. The two channel situation is with regard to the subsequent processing by the sequential tracking mechanism. The hypothesized increase in channel separation as a function of time is supported by the observed increase in gap detection thresholds for the longer duration tonal stimuli.

The argument for contour development, or "frequency tuning," as an integral process of the auditory nervous system was recently made by Elliot (1967). Support for the hypothesis came from masking data which revealed a strong degree of frequency selective masking over long intervals between tone probe and narrowband maskers. It was postulated that contours required approximately 250-300 msec. to completely develop, with reference made to frequencies in the region above 1000 Hz. A slower rate of contour development was suggested for the much lower frequencies, the region for which the auditory system is believed to rely less on place of cochlear excitation for its pitch information. This latter assumption, if valid, would explain the observed similarity in "continuity" thresholds (discussed above) for low frequency and noise streams. Also, if simple frequency contours develop completely after a certain time interval, as claimed, one might expect to observe a concomitant asymptote in gap thresholds for continued increases in tonal duration, with fixed frequency setting. Unfortunately, available gap detection data was obtained with stimuli not greater than 300 msec.

That even partially developed frequency contours may function as "pre-tuners" is also supported by experimental findings. Shipley (1959) demonstrated that weak signals are better detected when preceded by stronger ones of the same frequency. Recently, Ronkin (1972) observed similar effects in a task which required listeners to choose one of two temporal intervals containing the higher pitched 10 msec. signal. The best performance was observed in conditions when a 40 msec. tuning signal of similar frequency immediately preceded the probe burst. If the irrelevant signal followed the probe, no improvement over the probe alone condition was observed. It would seem reasonable to suggest that the irrelevant stimulus, and its partially developed contour representation, served as a reference point in the frequency space from which to estimate contour separation, or which may have incorporated the probe information to modify its own contour in some discriminable manner.

Although it is difficult to generalize from simple frequency contours to more complex structures, the analogy between the perceptual integrity of tonal streams and simple contour development seems most direct. Sequential patterns of alternating tonal stimuli, characterized by wide jumps in frequency, reveal little concerning the developmental aspects of perceptual streams. Even a gross initial sorting of inputs by frequency domain would be expected to prevent unitary perceptual patterns from occurring. For other less frequency variable patterns, however, it was mentioned previously that singular perceptual structures eventually split with an increase in either presentation rate or listening duration. Such steady but intermittent stimulation of the same sensory input locations might be expected to result in the same "priming" function as an increase in duration for single tonal stimuli. Eventually in this process, the units for perceptual selection would be modified as melodic continuity broke down and multiple frequency contours developed. With regard to speech stimuli, the same process may be found to underlie the "verbal transformation effect" reported by Warren (1968). He demonstrated that if a single word or short phrase is presented in a continuously repetitive manner, listeners experienced perceptual transformations. The word, "rest," for example, was perceptually changed to "stress," "tress," or "Esther." Similar reversals of structure were reported by listeners in the Thomas _et al._ (1970) study which employed repeated presentation of vowel sequences. That perceptual transformations for these "unnatural" speech conditions were reported as meaningful words may reflect the constructive listening processes employed by individuals. It seems reasonable to assume, however, that the perceptual units which were utilized in the process of semantic synthesis were developed preattentively, as phonemic contours developed through repetitive stimulation.

At the present time, too little is certain concerning the nature of simple contours to allow meaningful discussion of the more complex frequency structures. The combined effects of frequency, duration and switch time seem to explain listener performance in the detection of brief temporal gaps between elements of simple auditory sequences. More information needs to be gathered concerning contour development times as a function of frequency, and the interaction of successive contours in the perceptual development of entities displaying Gestalten qualities. Any discussion of critical frequency bands would seem more appropriate to investigation of minimum non-interactive separations for fully developed, simple frequency contours. For more complex auditory patterns, it is also possible that continued tuning occurs, beyond presence of acoustical energy, in the uniting of simple contours. Ultimately, one might wonder if a simple function relating switch time and switch "distance" remains applicable. The interpretation of switching attention to an element of a complex structure might well be complicated by the finding that such focal shifts are directed to the whole structure initially, before the processing of parts -- the latter stage itself a time consuming process.

REFERENCES

Abel, S. M., "Duration Discrimination of Noise and Tone Bursts," Journal of the Acoustical Society of America, 1972, 51, 1219-1223

Bregman, A. L. and Achim, A., "Visual Stream Segregation: Splitting of the 'phi,'' Paper presented at Eastern Psychological Association meeting, Boston, April, 1972

Bregman, A. S. and Campbell, J., "Primary Auditory Stream Segregation and Perception of Order in Rapid Sequences of Tones," Journal of Experimental Psychology, 1971, 89, 244-249

Broadbent, D. E., "Mechanical Model for Human Attention and Immediate Memory," Psychological Review, 1957b, 64, 205-215

Broadbent, D. E., "Immediate Memory and Simultaneous Stimuli," Quarterly Journal of Experimental Psychology, 1957a, 9, 1-11

Broadbent, D. E., Perception and Communication, New York, Pergamon Press, 1958

Broadbent, D. E., "The Role of Auditory Localization and Attention in Memory Span," Journal of Experimental Psychology, 1954, 47, 191-196

Broadbent, D. E., "Successive Responses to Simultaneous Stimuli," Quarterly Journal of Experimental Psychology, 1956, 8, 145-152

Broadbent, D.E. and Gregory, M., "On the Recall of Stimuli Presented Alternately to Two Sense Organs," Quarterly Journal of Experimental Psychology, 1961, 13, 103-109

Bukofzer, M., Music in the Baroque Era, New York, W. W. Norton, 1947

Chandler, A. R., Beauty and Human Nature, New York, Appleton-Century-Crofts, 1934

Day, R. S., Fusion in Dichotic Listening, Unpublished doctoral dissertation, 1968, Stanford University

Deutch, J. and Deutch, D., "Attention: Some Theoretical Considerations," Psychological Review, 1963, 70, 80-90

Elfner, L. F. and Caskey, W. E., "Continuity Effects with Alternately Sounded Noise and Tone Signals as a Function of Manner of Presentation," *Journal of the Acoustical Society of America*, 1965, *38*, 543-547

Elliot, L. L., "Backward and Forward Masking of Probe Tones of Different Frequencies," *Journal of the Acoustical Society of America*, 1962, *34*, 1116-1117

Elliot, L. L., "Backward Masking: Different Durations of the Masking Stimulus," *Journal of the Acoustical Society of America*, 1964, *36*, 393 (L)

Elliot, L. L., "Development of Auditory Narrowband Frequency Contours," *Journal of the Acoustical Society of America*, 1967, *42*, 143-153

Foder, J. A. and Bever, T. G., "The Psychological Reality of Linguistic Segments," *Journal of Verbal Learning and Behavior*, 1965, *4*, 414-420

Gray, G. W., "Phonemic Microtomy: The Minimum Duration of Perceptible Speech Sounds," *Speech Monographs*, IX, 1942, 75-90

Heise, G. A. and Miller, G. A., "An Experimental Study of Auditory Patterns," *American Journal of Psychology*, 1951, *64*, 68-77

Hirsh, I. J., "Auditory Perception of Temporal Order," *Journal of the Acoustical Society of America*, 1957, *31*, 759-767

Homick, J. L. and Elfner, L. F. and Bothe, G. G., "Auditory Temporal Masking and the Perception of Order," *Journal of the Acoustical Society of America*, 1969, *45*, 712-718

House, A. S., "On Vowel Duration in English," *Journal of the Acoustical Society of America*, 1961, *33*, 1174-1178

Ladefoged, P. and Broadbent, D. E., "Perception of Sequence in Auditory Events," *Quarterly Journal of Experimental Psychology*, 1960, *12*, 162-170

Miller, G. A. and Heise, G. A., "The Trill Threshold," *Journal of the Acoustical Society of America*, 1950, *22*, 637-638

Moray, N., *Attention: Selective Processes in Vision and Hearing*, London, Hutchinson Educational Ltd., 1969

Moray, N., "Broadbent's Filter Theory: Postulate H and the Problem of Switching Time," *Quarterly Journal of Experimental Psychology*, 1960, *12*, 214-220

Neisser, U., *Cognitive Psychology*, Appleton-Century-Crofts, 1967

Norman, D. A., "Temporal Confusions and Limited Capacity Processors," *Acta Psychologica*, 1967, *27*, 293-297

Ortmann, O., "On the Melodic Relativity of Tones," *Psychological Monographs*, 1926, *35* (1, whole No. 162)

Plomp, R., "Rate of Decay of Auditory Sensation," *Journal of the Acoustical Society of America*, 1964, *36*, 277-282

Reber, A. S. and Anderson, J. R., "The Perception of Clicks in Linguistic and Nonlinguistic Messages," *Perception and Psychophysics*, 1970, *8*, 81-89

Ronkin, D. A., "Changes in Frequency Discrimination Caused by Leading and Trailing Tones," *Journal of the Acoustical Society of America*, 1972, *51*, 1947-1950

Shipley, E. F., "Cueing as a Determiner of Apparent Variability in Sensitivity," *Journal of the Acoustical Society of America*, 1959, *31*, 834 (abstract)

Small, A. M., and Campbell, R. A., "Temporal Differential Sensitivity for Auditory Stimuli," *American Journal of Psychology*, 1961, *75*, 401-410

Swets, J. A., "Central Factors in Auditory Frequency Selectivity," *Psychological Bulletin*, 1963, 429-440

Thomas, I. B. and Fitzgibbons, P. J., "Temporal Order and Perceptual Classes," *Journal of the Acoustical Society of America*, 1971 (A), *50*, 86-87

Thomas, I. B., Hill, P. B., Carroll, F. S. and Bienvenido, G., "Temporal Order in the Perception of Vowels," *Journal of the Acoustical Society of America*, 1970, *48*, 1010-1013

Thurlow, W. R. and Elfner, L. F., "Continuity Effects with Alternately Sounding Tones," *Journal of the Acoustical Society of America*, 1959, *31*, 1337-1339

Treisman, A. M., "The Effect of Irrelevant Material on the Efficiency of Selective Listening," *American Journal of Psychology*, 1964, *77*, 533-546 von Bekesy, Experiments in Hearing, New York, McGraw-Hill, 1960

Warren, R. M., "Relation of Verbal Transformations to Other Perceptual Phenomena," Paper presented at IEE/NPL Conference on Pattern Recognition, Teddington, England, July, 1968

Warren, R. M., Obusek, C. J., Farmer, R. M. and Warren, R. P., "Auditory Sequences: Confusions of Patterns Other Than Speech or Music," Science, 1969, 164, 586-587

Williams, K. N. and Perrott, D. R., "Temporal Resolution of Tonal Pulses," Journal of the Acoustical Society of America, 1972, 51, 644-647