1990

Perceptual correlates of spectral changes in complex tones

Punita Gurpreet Singh

WASHINGTON UNIVERSITY

Department of Speech and Hearing
Committee on Communication Sciences

Dissertation Committee:

Ira J. Hirsh, Chairperson
James D. Miller
Julius L. Goldstein

PERCEPTUAL CORRELATES OF SPECTRAL CHANGES IN COMPLEX TONES

by

Punita Gurpreet Singh

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfilment of the
requirements for the degree
of Doctor of Philosophy

December 1990

Saint Louis, Missouri

"SENSATIONS OF TONE"

still awesome, still sensational
Some solved, some "resolved" components,
some absolute, some relational

Dissonance and brightness
tone height and tone chroma
The pitch of the "residue"
a "fundamental" misnomer

The partials of a complex,
what causes them to fuse ?
what facilitates their being grouped together,
what are the underlying perceptual cues ?

What enables the segregation
of an entangled acoustic compound
into separate perceptual entities,
each with its own special sound ?

Is it frequency, is it intensity,
is it a spectro-temporal difference ?
Or a combination of factors,
that enable this inference ?

From the time of von Helmholtz,
there has been a recurring hypothesis
of a two-stage process
with synthesis following analysis

A peripheral filtering,
a spectral sorting mechanism
Or a temporal feature matching,
Maybe both, . . . there's still some skepticism

A central putting together,
of bits of information
Noting patterns and correlations
that aid "image" formation

The input sorted out,
individual sources defined
Their pitch and timbre and location
determined and assigned

But this complacent scenario
is disturbed by what's heard next
a sudden change in meaning
provided by a new context

Dynamic and flexible
ongoing comparisons being made
Simultaneous and sequential
incoming data being weighed

Signal detection
in a morass of sound
a release from masking,
a "figure" emerging from the "ground"

The forming of "gestalts"
guided by logical laws
a seemingly simple relation
between effect and cause

But the processing operations
are difficult to elucidate
To measure and to quantify,
to model and to validate

What was believed found is sometimes lost,
and answers left in the lurch
But the endeavor ends never
and research continues to re-search

Fundamentally,
Ms. Singh, 1989

ABSTRACT

The traditionally maintained separateness of "timbre" from other tonal percepts such as "pitch" is questioned in a set of experiments designed to ascertain perceptual cues facilitating auditory discrimination tasks. Three types of complex sounds with flat spectral envelopes are studied: 1) harmonic "residue" tones comprising 4 harmonics, 2)10-component harmonic and inharmonic complexes with all components shifted from some reference frequency, and 3)10-component complexes with a single component shifted from its harmonic frequency. With 2-tone sequences used as stimuli, listeners are asked to judge if the second tone of a pair is 1) same, 2) higher in pitch, 3) lower in pitch, 4) different in "something else", 5) different in "something else" and higher, or 6) different in "something else" and lower in pitch than the first tone. ("Something else" is taken to be synonymous with "timbre"). For residue tones, the data indicate that changes in spectral locus yield changes in timbre. Further, the direction of locus change can indicate a pitch change, despite little or no change in fundamental frequency (0-2% for 200 and 400 Hz F0). This implies that a change in timbral "sharpness" may be construed as a change in pitch, given the absence of other cues.

For stimuli in which one or more components are shifted from harmonic frequencies, the unitary sensation of a complex may be replaced by one of multiple sources. A second task enabled reports of

such perceptual "fission". Low components are more susceptible to being "heard out" as individual entities, while changes in higher components may yield changes in timbre, such as "roughness". Further, context-dependencies are indicated, with adjacent components being compared both within and across sounds for judgments of **fusion, pitch** and **timbre** change.

Thus, local and global comparisons of components based on factors such as: 1) **magnitude**, 2) **location**, 3) **direction**, and 4) **context of spectral change** in a sequence, guide basic grouping operations, in addition to indicating changes in an overall property of a complex sound as a whole, such as its pitch, timbre, or both of these percepts simultaneously.

PREFACE and ACKNOWLEDGMENTS

The completion of this doctoral dissertation represents the formal consummation of a lifelong desire to be an "acoustician".

A childhood fascination with music was noticed and nurtured by parents who provided all kinds of interesting instruments for play and exploration . At age fourteen, an introduction to "harmonics" in a physics class, led to a change in my career ambition, from wanting to be a musician, to wanting to be a musician *and* a scientist. After undergraduate years spent studying physics, serendipity guided me to an interdisciplinary master's program in psychoacoustics that proved to be an incredible, ear-opening experience. Given the freedom to pursue courses in music, architecture, biology, engineering, speech and hearing, physics and psychology, my many questions about creating, storing, transmitting and perceiving sound received attention.

I am indebted to Drs. William H. Danforth, Ira Hirsh, Dick Norberg, the late Tom Sandel, Dean Edward Wilson, and others who collaborated in letting me be a part of such an educational renaissance adventure.

After an all-consuming thesis project on "dimensional tradeoffs" conducted at the laboratories of Central Institute for the Deaf, I stayed on to continue research as a doctoral student in "Communication Sciences".

Alas, there followed a tumultuous period in which something traumatic seemed to happen every day . . .

Even as I was studying factors involved in grouping and segregation of sounds, my beloved homeland was being ripped apart by analogous issues of separatism and communal dissonance. The delight and wonder of life were dampened in the face of all the disharmony resonating in the world. Other personal losses and disappointments led to a misanthropic depression that has been difficult to surmount at times. But the love, care and optimism of family and friends, and continued excitement of the world of sound have been strengthening and uplifting through all crises.

I wish I could personally thank and repay all who helped me endure this rough period. Perhaps the best way to express my gratitude is to resolve never to succumb to such depths again and to carry on the business of living with joy and zest and faith.

The research carried out for this doctoral project received financial, intellectual and emotional support from many quarters:

The United States Air Force funded the work for the most part under grant AFOSR - 87-0382. In this last year, financial support was also made available through an NIH grant to Central Institute (BRSG - S07 RR05987).

To my mentor Dr. Ira Hirsh I am grateful for the wisdom, humor, and incredible patience, with which he has handled me through my transitions, digressions and regression from being a mature, stable *"etudiante modele"* - to being a fluctuant *"enfant terrible"*. The

consistent support and affection I have received from him and from Shirley Hirsh have carried me through many tough times.

To my dissertation committee, I am grateful for the insights, *saykhel* and suggestions they have shared with me. Dr. Julius Goldstein devoted considerable time to critiquing the work and pointing out important directions in which it could be developed and improved. Dr. James D. Miller has shown periodic enthusiasm for my work since the master's project on "streaming". Dr. Judith Lauter continues to serve as a model for successful integration of diverse interests and provided encouragement, advice and affection. The marathon reading effort by Drs. William Clark, Michael Friedlander, Roland Jordan and Edward Wilson, and their comments are also appreciated. I feel fortunate to have had such a strong and diverse readership on my examining committee.

My maturation as an acoustician has also been directly influenced by reading, and meeting, Drs. Arthur Benade, Albert Bregman, Pierre Divenyi, William Hartmann, Adrian Houtsma, Mari Reiss Jones, Steve McAdams, Reinier Plomp, Ernst Terhardt, and David Wessel . A summer spent in Art Benade's lab working on issues related to the preservation of timbral identity despite variation of spectra in rooms proved to be the launching pad for research on timbre perception.

For providing *in vivo* acoustical training, I am grateful to my drum teachers - Zameer Ahmed and Monduel Banessia, and to Ms. Sheila Dhar, Rabia Sangwan, David Hykes and Rich O' Donnell, all of whom bring to life the principles of organizing sound I have sought to study.

Many colleagues at CID helped me in my efforts to complete this work: JoAnne Kocunik played a major role, in providing ears, laboratory assistance, help with graphics, and a general cheery presence in and out of the lab. Dr. Robert Gilkey has been generous with time for discussions and use of computing facilities. Kit Lai and Dr. Maynard Engebretson provided software support. Arnold Heidebreder, Bill Miksicek, Steve Sadoff, Michael O'Connell, and Clarence Rulo helped out with technical problems. Kathy Eckenrod zealously monitored my staying on (and straying from) the proper academic-administrative path. Mary Sicking ensured access to valuable reading materials. Many others lent ears and shoulders and gave advice from time to time - Carole Campbell, Sally Charton, Marios Fourakis, Marilyn French-St. George, Michael Gottfried, Ken Grant, John Hawks, Dave Hillier, Caroline Monahan, Johanna Nicholas, David Pascoe, Lourdes Peironcely, Don Ronken, Marty Silverman, Brian Simpson, Richard Stoker, Leigh Tenkku, Jan Weisenberger, and the folks down in "SDL" and "TPL".

To my listeners, Carole, Cynthia, George, Hyla, JoAnne, John, Judi, Lara, Lee, Lori and Wendy, - I am grateful for their cheerful endurance of hours of my tones and groans and for their articulate, revealing comments about the stimuli.

My long and multi-faceted stay at Washington University has fostered the forging of many friendships and professional alliances.

The Music Department has been a special haven through the years, - providing employment, enjoyment, a forum to teach, to learn and to remain in close touch with musicians and with the original musical issues that brought me here. The Department of Residential Life too, has been a special place, - providing a home, friends, and the training to handle crises without panic !

My own stresses and strains were eased by the additional support of friends - Rajiv, Sanjiv and Goody Chhatwal, Arun and Abha Kumar, Wendy and the Katz *mishpawkhuh,* Jack and Sarah Burke, John Arnold, Chris Loving, Marc Weiner, and the long distance affection and counseling provided by Madhav Dhar, Ashwini Kapoor, Joe and Jae Rand, Asha Singh , Marta and John Tetzeli, Meeti, Geeta and Devika.

The longest-lasting, long-distance support of all has been provided by my incredible family. My Father nursed me out of many a slump with his soothing wisdom, dynamism, and strength. My Mother uprooted herself from home ground for a while, to assist me in the birth of this dissertation baby - labor that turned out to be pleasurable and bonding for both of us ! My siblings Ina and Simi and their couplings Reena, Sanjeev, Tara, Amira and little Prithvi shared the ups and downs and consistently provided affection, care, energy, and humor.

With all the uncertainties and insecurities around, it's reassuring to know that some relations endure, despite time, distance, distorting nonlinearities and changed frames of reference.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter One

FOREGROUND

1.1 Aim and motivation

The research described in this dissertation is part of an ongoing effort to understand and investigate the physical and psychological principles active in the creation and **perception of music** (Singh, 1984, 1985, 1987, 1988, 1989).

Acoustic systems of communication and expression such as speech and music reflect the ability of the auditory system to organize sound perceptually. **"Organization"** inherently implies some type of structuring or ordering of events. The ability to organize sounds simultaneously and sequentially in time is of primary importance in music. The development of vertical ("chordal") sound structures in **harmony**, and the horizontal, temporal structures of **melody** reflect such organization. Harmony and melody are two closely related **"shaping forces"** in music (Toch, 1948/1977). "While harmony is marked by the temporal coincidence of different pitches, melody is marked by their temporal succession" (p. 62).

The perception of relations is basic to the melodic and harmonic organization of sounds in music. Both melodies and chords can be considered to be **groups of musical intervals** that are defined by ***relations*** between the notes bounding the intervals, such as the ratios

of their fundamental frequencies. The evolution of these structures over time unfolds changing relations between notes. The very existence of music implies that the auditory system is capable of following these changes and analyzing and reinterpreting the acoustic "scene" dynamically (Bregman, 1990).

Many contextual factors influence the organization of sounds. Factors such as pitch proximity, rate of presentation (temporal proximity) and timbre similarity can all serve as cues guiding the sorting of sound sequences into appropriate groups. The study of **sequences** has mostly been approached from a **psycho-*musical*** viewpoint where the **"note"** is the elemental unit of sound with an overall perceptual attribute such as pitch, timbre, loudness, and duration. Relations between such perceptual attributes of sounds and the influence of the time dimension embedding the sequence have been the focus of some studies (Balzano, 1986; Bregman, 1990; Erickson, 1975; Jones, 1976; Krumhansl, 1979; Monahan and Carterette, 1985).

The study of different tonal features themselves, on the other hand, has usually been approached from a more **psycho-*acoustical*** point of view (Goldstein, 1973; Plomp, 1970, 1976; Zwicker and Scharf, 1965). Such studies typically treat a "note" as a complex sound **"event"** comprising an aggregate of simpler "spectral" components. The contribution of these more elementary acoustic units to the perception of overall attributes such as pitch, timbre and loudness, and the limits of sensory processing are often the focus of investigation, divorced from

musical considerations of how these features are perceived in relation to each other.

However, the perception of relations is also important at the event level for feature-extraction operations that enable unitary percepts such as pitch and phonemic identity to be assigned to *groups* of spectral components. For the perception of pitch, the important relations may be, for example, the harmonicity of components that render them multiples of a common fundamental frequency (Goldstein, 1973). For the perception of phonemic identity, an example of relation perception is the comparison of formant-frequency ratios to yield different vowel sounds (Fant, 1973; Miller, 1984, 1988).

The experiments conducted for this dissertation stand on middle ground given the two investigative approaches mentioned above. The underlying questions that motivated the research have been concerned from the outset, both with determining the physical bases of perceptual attributes and with the understanding of relations between such features of sounds in sequence.

The perceptual grouping of **sequences** was studied in an earlier experiment that evolved to yield the present experiments (Singh, 1984/1987). The organization of a sequence of sound events may be influenced by relations between perceptual features assigned to the individual events. The process of sequencing, often augmented by repetition, allows similarities and differences between sounds to be discovered and used as criteria in organization and categorization. Given

the dynamic, changing context of a sequence, a listener is faced not only with the task of observing features of individual events, such as pitch, timbre and loudness, but also the task of tracking changes along these dimensions from event to event over time. Both spectral and temporal relations between events may influence how a sequence will be perceptually organized.

Temporal relations between auditory events in a sequence can guide its perceptual organization (Hirsh, 1974). The onset-to-onset time intervals separating sound events influence discriminability of their order of occurrence in the sequence (Hirsh,1959). While accuracy in ordering prevails at slow rates of presentation, quicker rates are seen to result in confusions in the perception of order. Furthermore, the rate at which this occurs, varies, depending on characteristics of the sounds comprising the sequence (Warren et al.,1969).

Sounds that differ greatly from one another may appear to break apart at quick rates to form subsequences, within which the range of differences is smaller. This perceptual segregation of a sequence has been referred to as "fission" (Dowling, 1968; van Noorden, 1971) and **"stream segregation"** (Bregman and Campbell, 1971). While it is easy to order events within a perceptual stream, it becomes difficult to judge their order across streams (Bregman and Campbell, 1971; Dannenbring and Bregman, 1976).

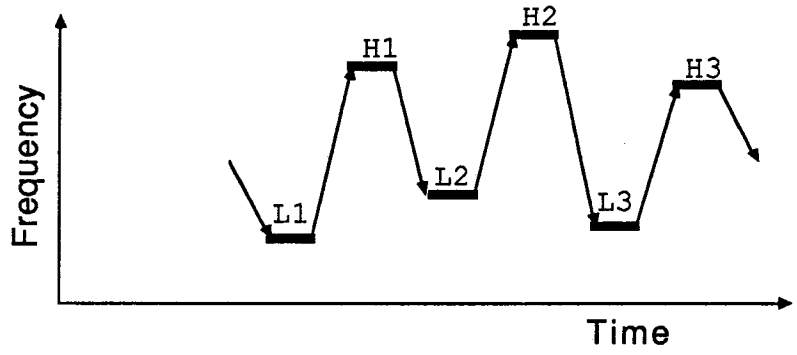
The reorganization into "streams" appears to be dependent on "lawful relations" between the perceptual dimensions of the sounds, such

as their relative pitch, loudness and duration and the time layout of the sequence in which they are embedded (Jones, 1976). These lawful relations may derive from **"gestalt"** principles of perceptual organization, such as 'common fate', 'good continuation', 'belongingness', and 'covariance' of features of the events (Koffka, 1935).

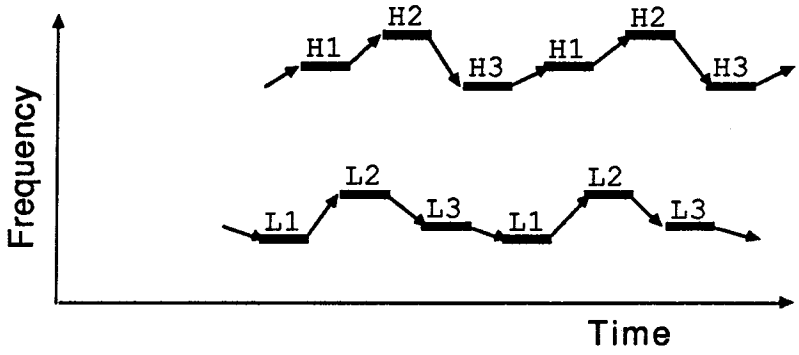
Figure 1.1 illustrates the streaming concept. The vertical axis represents frequency and the horizontal axis represents time. The six interleaved tones shown in the sequence belong to two distinctly different frequency regions, one being high (H) and the other low (L). When the tempo of the sequence is sufficiently rapid (e.g. 10 tones/sec), the sequence segregates perceptually into two streams, one comprising the high tones (H1, H2, H3) and the other the low tones (L1, L2, L3). Perception of order is accurate within a stream, but poor across the elements of the two streams.

The physical differences influencing organization are manifested as different perceptual criteria. In the example above, the sequence comprised a succession of pure tones of different frequencies. For this type of stimulus, the organizing cue is **proximity of physical frequency**, which is construed in terms of perceived pitch differences.

For sequences of complex tones, both pitch proximity and timbre similarity, have been shown to be effective organizing cues dictating stream segregation (McAdams, 1977; van Noorden, 1975; Singh, 1984, 1987; Wessel, 1979).



(a) One Stream



(b) Two Streams

Figure 1.1 A repeating 6-tone sequence of the type used by Bregman and Campbell (1971). In (a), the high (H) and low (L) tones alternate at a tempo of 5 tones/sec and a single "stream" is perceived. In (b), the tempo is twice as fast (10 tones/sec). The high tones segregate perceptually from the low tones and two overlapping streams are perceived. [Adapted from McAdams and Bregman, 1979].

The **role of timbre in stream segregation** was explored further in previous work by the author (Singh, 1984/1987). Harmonic "residue" tones were used as stimuli in that study. For such tones, the frequency spacing of spectral components and their position (or **"locus"**) along the frequency axis are generally correlated with pitch and timbre respectively (Schouten et al., 1962). Timbre differences were thus provided by varying the spectral locus of 4 equi-amplitude harmonics, and pitch differences were provided by varying their relative spacing. The pitch and timbre attributes were put into competition with each other in sequences of the type:

T2P1 -- TmP1 -- T2Pn -- TmPn

with the first pair of tones assigned pitch P1, but different timbres T2 and Tm, and the second pair assigned pitch Pn, and similarly- contrasted timbres. A pitch-based grouping of such a sequence would result in "veridical" perception with correct judgments of temporal order. A timbre-based grouping would yield incorrect judgments of order, with alternate tones of the same assigned timbre being grouped together.

Listeners were asked to indicate whether perceived grouping of such sequences was based on pitch proximity, timbre similarity, or ambiguous percepts not dominated by any cue, by selecting one of the response alternatives illustrated in **figure 1.2.**

The selection of the response label '2' for several of the timbres contrasted, confirmed that timbre can segregate sequences into streams. The distribution of the response labels further implied that **pitch and**

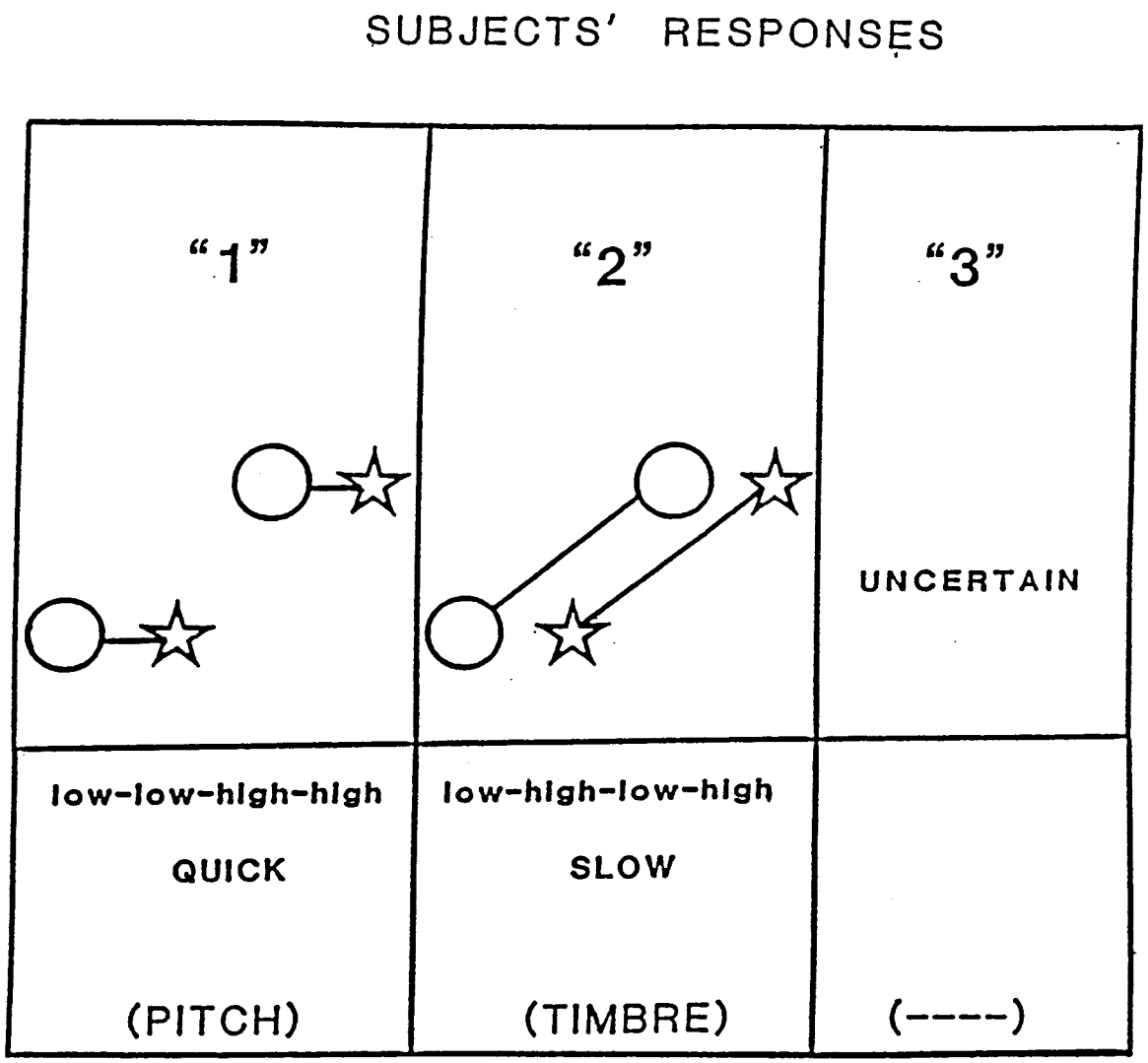


Figure 1.2 Response labels provided to listeners by Singh (1984) to indicate perceptual grouping of 4-tone sequences. Label 1 was used to denote a pitch-based grouping, label 2, a timbre-based grouping, and label 3, uncertainty in response. These groupings were confounded with the perceived tempo of the sequences. Due to disruption in temporal order perception, grouping 2 appeared to be slow (half the tempo of grouping 1, which appeared to be quick). [From Singh, 1987].

Since the **frequency spectrum contributes to both the pitch and the timbre of a sound**, spectral changes may be construed as changes in either pitch, or timbre, or both of these percepts simultaneously. These percepts can thus "compete" as cues guiding alternative organizations (Bregman, 1978). The observed interaction of pitch and timbre in the earlier work on "streaming" and the confounding influence of spectral locus on both pitch and timbre called for finer scrutiny of the stimuli used and the type of percepts associated with changes in spectral locus and spectral spacing. This investigation has been undertaken in this doctoral dissertation. The experiments reported herein were designed to focus **explicitly** on the issue of **perceptual correlates of spectral changes in complex tones.**

Instead of using long, fast, repeating sequences typical of the "streaming" paradigm, the simplest case of comparing **two tones** was studied. The first tone was used as a "standard" with which to compare a second tone. The second tone differed from the first in that one or more of its components were systematically changed in frequency. A specially-designed labelling task enabled listeners to report the percepts associated with different types of spectral changes.

1.2 The discrimination problem

When a sequence of two different sounds is presented to a listener, a variety of observations can allow the listener to discern that the sounds

are dissimilar. One sound may appear to be louder or longer in duration than the other. It may seem higher or lower in pitch. It may sound "rougher" or more "dense" or "brighter" than the other sound. Or there may be a combination of such features that aid the listener in making a discrimination judgement.

These sensations are related to physical characteristics of the presented sounds, but their relation is not always transparent. Psychophysical procedures attempt to determine the nature of such relations between psychological percepts and the underlying physical structures from which they are derived.

Thus, the psychological percept "pitch" has been observed to be related primarily to the physical feature "frequency of vibration". The perceptual attribute "loudness" has been observed to be related primarily to the physical features "intensity" and "sound pressure level". The perceptual attribute "timbre", however, continues to elude easily specifiable relations to physical structure. It is a "multidimensional" attribute that appears to be affected by both temporal features of sounds, and "spectral" features related to the amplitude distribution of frequencies present in the sound. To further complicate matters, in many situations, interactions are observed within and between the set of physical features of sounds, and their associated percepts.

Percepts are difficult to quantify in a numerical sense. Perception has been defined as being "essentially, the reading of meanings from sensory signals" (Harre and Lamb, 1983). While it has been suggested

that the study of human communication requires a bridging of the gap between purely physical aspects of theories concerning measurements on signals, and human aspects of communication concerning "meaning" and "effectiveness" (Gerson and Godstein, 1978 (p. 509); Shannon and Weaver, 1949), a vast amount of research in psychoacoustics has tried to circumvent the problem of specifying percepts directly by resorting instead to measurement of **sensory** limits. Such a limit may be, for example, the minimal change needed along some dimension such as intensity, duration, or frequency, that would enable **detecting** the presence of a sound, or **discriminating** a sound from a "standard" comparison sound, or lead to the correct **identification** of a sound. The measure of performance is usually taken to be that **"threshold"** value of the stimulus dimension being varied, that leads to a certain percent correct (e.g. 75%) in judgment. For discrimination tasks, this "threshold" is referred to as the "just noticeable difference" (jnd) or the "difference limen" (DL).

The response measure employed in **frequency discrimination** tasks is thus typically a measure of the limiting value of frequency change that allows a listener to discriminate between two sounds. Historically, in the vast literature on frequency discrimination, one mostly encounters studies that used pure-tone signals as stimuli (e.g. Harris, 1966; Moore, 1973; Shower and Biddulph, 1931; Weir et al., 1977). These "simple" signals comprise a single frequency of vibration and frequency discrimination is thus congruent to pitch discrimination.

An upward change in frequency is typically perceived as a rise in pitch, while a downward change is perceived as a fall in pitch.

"Complex" sounds on the other hand, comprise aggregates of simple, "partial", spectral components (ASA, 1960). Changes in frequency can therefore potentially be brought about in a multitude of ways. Changes can be made in the fundamental frequency (F0) of a harmonic complex tone, or in the frequency of one or more components in the spectrum. For broadband signals, the bandwidth or the center frequency of the spectral distribution may be changed. Similar changes may be made for resonance frequency (peaks or "formants" in the spectrum). Such changes can influence the pitch percept, but can as well result in qualitative, timbral differences. All these spectral changes could be studied in detection or discrimination paradigms and values obtained for frequency shifts required at threshold, without obtaining any information about **how or why** a listener came to decide on a particular judgment.

1.3 Discrimination of spectral changes in complex sounds

Over the years, many experiments have been conducted that called for spectral discrimination judgments. Experiments on frequency discrimination, vowel-formant discrimination, auditory masking, profile analysis etc. all require **comparisons of spectra**. The measure of performance is usually the magnitude of the changed dimension needed

to reach some level of accuracy. The listener is often required simply to state, for example, if a pair of stimuli are same or different, or to identify the presentation interval that contains the "changed" stimulus. While the results give important information about thresholds of discriminability, they do not reveal the **perceptual cues** used by listeners in making their judgments.

The experimental tasks employed in this dissertation project were specifically designed to find out what these perceptual cues might be. In particular, the perceptual correlates of changes made in the frequency of one or more components of a complex tone were the major focus of the investigation.

This territory is not completely uncharted. McAdams (1984 b), Moore et al. (1984, 1985, 1986), and Hartmann (1988) have used stimuli in which components of a tonal complex were displaced from some reference frequency. Tasks for their experiments ranged from judgments of pitch change and perceived "mistuning", to judgments of splitting of a sound into multiple sources or "entities".

All these percepts can serve as **potential cues** for spectral discrimination tasks. The experiments reported here sought to **map out** this range of percepts and their relation to factors such as the **magnitude** and **location** of spectral changes, and whether these left the complex **inharmonic,** or **spectrally dislocated** along the frequency axis.

1.4 Overview of experiments

On any given "trial" in an experiment, listeners were presented a sequence of two complex sounds that were either the same, or differed in the frequency of one or more spectral components. In all cases, the stimuli were designed to have identical durations, temporal onset and offset characteristics, and equal amplitude of components. The only dimension being changed was frequency. The aim of the experiments was to obtain information about the perceptual criteria guiding discrimination judgments, not to determine "thresholds" of discrimination.

While changes in the frequency of spectral components often yield changes in pitch, the perceived change may equally often be in some qualitative feature such as "roughness" or "brightness". Listeners were thus given a number of **different options** to report their percepts. These included, and extended, the traditional "same"/"different" option provided in many frequency discrimination experiments. Thus, one of the tasks in the experiments asked listeners to report if the second sound of the pair was 1) same, 2) higher in pitch, 3) lower in pitch, 4) different in "something else" but same in pitch, 5) different in "something else" and higher in pitch or 6) different in "something else" and lower in pitch than the first sound.

The seemingly-ambiguous term **"something else"** was used intentionally to cover those timbral percepts that elude ease of

description and is taken to be **synonymous with "timbre"** (which was described to listeners as being a feature of sounds distinct from their pitch, loudness or duration).

In some cases, changes in the frequency of spectral components can bring into play **segregation processes** that change the unitary sensation of a complex sound into one of multiple sources (Cohen, 1980; McAdams, 1984 b; Moore, 1986; Hartmann, 1988). This is observed to happen primarily for those stimuli in which changes in frequency of components lead to the sound becoming "inharmonic". The partial components of such stimuli are not related in a simple numerical fashion as are the partials of a "harmonic" complex. In the latter, all are integral multiples of a common, "fundamental" frequency (F0). The components of inharmonic stimuli do not share this redundancy feature of being multiples of a common denominator. For such stimuli, an additional task requested listeners to report if either of the two sounds of a stimulus sequence appeared to "split" or segregate into more than one entity.

Three related experiments were conducted, with the listener assigned the multiple-option tasks described above. The complex sounds employed as stimuli differed in terms of the **density, harmonicity,** and **location of spectra**. The options that were selected most frequently in labelling particular stimuli revealed the perceptual criteria that guided discrimination.

Results show that listeners are able to make **context- dependent comparisons of spectra.** Big jumps in spectral location, while usually

correlated with timbre change, can be confused with changes in pitch. Inharmonic changes may result in multiple perceptual changes. Small shifts in frequency contribute to changes in pitch. Larger changes, when made in the frequency of low components typically lead to segregation, and for high components, to changes in timbre. The design of the stimuli and task employed in the three experiments thus enabled the perceptual correlates of spectral changes to be inferred from listeners' responses.

1.5 Summary of issues addressed by dissertation

The research documented in this "bi-focal" dissertation is an endeavor toward furthering our understanding of the processes that guide our ability to differentiate components of a compound acoustic environment into perceptual units and to assign them features such as "pitch" and "timbre". The relation of these percepts to the underlying physical structure of the sounds is one focus of the work. The inter-sound context of the acoustic information available to the listener and its influence on the determination of sources and their perceptual features is another focus.

The three experiments reported in this dissertation cumulatively address the following related questions:

1. What are the perceptual cues underlying the discrimination of frequency changes made in one or more components of complex tones?

2. How do perceptual attributes such as pitch and timbre map on

to spectral features such as spacing and locus of components ? Do these percepts influence each other?

3. How do changes in the frequency of individual components affect their perceptual grouping? For cases where a complex appears to segregate into more than one entity, how are these entities separately perceived? What percept is associated with changes in frequency of single components for cases where the complex remains fused? What are the particular stimulus conditions that facilitate fission or fusion?

The format of the experiments designed to address these issues is **unique** in that it provides listeners with the opportunity to report percepts in a more explicit way than most experiments on spectral discrimination. This type of integration of tasks and their comparison for different types of stimuli has **not** been provided in most previous research. Such description then allows investigation of relations between physical changes made in the stimuli and in perceptual features such as pitch and timbre assigned to entities constructed from the multitude of components present in complex sounds.

1.6 Layout of Dissertation

The next chapter reviews extant knowledge about the physical structure of sound events, frequency sensitivity of the auditory system, fusion of components to form unitary "events", perception of event

features such as pitch and timbre, spectral discrimination, and the percepts associated with spectral changes. The review offers a broad perspective of the research areas. A more focussed and pertinent review of prior research is also given as needed in the individual "Introduction" and "Discussion" sections of chapters 3, 4 and 5 in which specific details of the experiments are described.

Chapter 6 offers a unified discussion of the results of all three experiments. The information gleaned about "perceptual correlates of spectral changes" is reviewed in the larger context of perceptual organization. The extraction of perceptual features from perceived auditory groups is contrasted with the extraction of auditory groups based on similarities of perceptual features. Finally, limitations of the present work and implications for future research are discussed in chapter 7.

Chapter 2
BACKGROUND

2.1 Chapter outline

This chapter constitutes a historical review and discussion of some solved and unsolved issues in research on auditory processing of complex sounds. The aim of the dissertation project is to ascertain perceptual correlates of spectral changes in complex tones. While spectral changes have been studied in many areas of research on hearing, these areas have largely maintained a distance from each other, given different scientific orientations, questions of interest, levels of investigation, and varied methodological approaches.

Experiments on frequency discrimination, signal detection, formant-frequency discrimination, pitch and timbre discrimination, all entail comparisons of spectra. This chapter attempts to review several of these areas and point out connections between them. Given the diversity and relevance of research on different aspects of complex sound processing, the material reviewed is somewhat expansive:

In **section 2.2**, complex sounds are first discussed in terms of the raw spectral material they are composed of. The ability of the ear to spectrally analyze complex sounds into simpler units is discussed next in **section 2.3.** However, in many listening situations, a "synthetic" mode is employed whereby complex sounds are heard as fused units with

overall properties such as pitch, timbre and loudness. Factors that facilitate the perceptual fusion of components to form a "gestalt" unit such as a complex tone are discussed in **section 2.4**, followed by a review of research on perceptual attributes such as timbre and pitch in **sections 2.5 through 2.8.** The common spectral basis of pitch and timbre and their potential interaction are discussed next in **section 2.9.** Experiments on "spectral" discrimination that highlight the variety of perceptual cues associated with frequency changes are then described in **section 2.10** as a segue to the main issues addressed by the dissertation, namely those pertaining to perceptual correlates of spectral changes. The aim of the experiments is restated in **section 2.11.**

The purpose of this lengthy review is to provide a repository of information pertinent to perceptual aspects of complex sounds and their discrimination and organization. For the sake of completeness, some historical details are included that may seem superfluous in light of current information. However, these are viewed as important in showing the evolution of ideas and the scientific treatment accorded over the years to different areas of perceptual processing of sound.

Hurried readers may scan the section headings and selectively read portions of the review of direct interest to them. The overview provided in chapter 1 and the introductory sections of chapters 3, 4 and 5 should be adequate for understanding the background and motivation for the experiments.

2.2 A "partial" view of complex sounds

2.2.1 Complex tones and the Fourier spectrum

Most natural sound sources generate sounds that are "complex". Such complexity reflects underlying oscillatory behavior that may be periodic (as for generation of tonal sounds) or aperiodic (as for generation of noise). The extension of a theorem by Fourier (1823) showed that a complex, periodic sound could be described mathematically, as being equivalent to a sum of simple sounds or "partials", that were multiples of a common, "fundamental" frequency (see Tolstov, 1962). The time function $f(t)$ representing a periodic signal with a period $1/f_0$, may then be expressed as a Fourier series expansion given by:

$$f(t) = a_0 + \sum_{n=1}^{\infty} (a_n \cos 2\pi n f_0 t + b_n \sin 2\pi n f_0 t)$$

where $f_0 = F0$ = fundamental frequency,

$n = 1 - \infty$ = component number,

a_0 = a constant term,

an, bn = amplitude factors.

In reality, the series will not usually comprise an infinite number of components. A complex tone may thus equivalently be represented by:

$$f(t) = \sum_{n=1}^{m} a_n \sin(2\pi n f_0 t + \phi_n)$$

[or by $f(t) = \sum_{n=1}^{m} a_n \cos(2\pi n f_0 t + \phi_n)$],

in which each component "n" (called a **"harmonic"**) is considered to be a "pure" tone, characterized by its own frequency (=nF_0), amplitude (=a_n) and phase (ϕ_n).

The distribution of frequency components and their amplitudes a_n comprise the **"Fourier spectrum"** of a sound. In most natural sounds, the partials fluctuate in frequency and amplitude over the duration of a sound. The true spectrum is therefore not a "static" feature. The initial, "attack" portion of a sound may be characterized by a different distribution of spectral components than the middle or the end portions. Most measures of the spectrum entail some sort of averaging procedure over time. The resultant time-averaged spectrum is a useful indicator of the general distribution of energy across frequencies and is used routinely as a descriptor of the physical parameters of sounds and their sources.

2.2.2 Ohm's acoustical law and the audibility of components of complex tones

The decomposition of a complex periodic sound into simpler, component part tones was considered by Ohm (1843) to be not just a mathematical equivalence, but something the ear was able to do in reality. His definition of "tone" (also known as "Ohm's acoustical law"), implies that a tone with frequency "f" can be heard if a complex sound

contains a simple "part tone" represented by the function $\sin(2\pi ft+\phi)$ as a component (see de Boer (1976, p. 492) for a more complete statement).

The auditory validity of Ohm's acoustical law was empirically tested by von Helmholtz (1877/1954). He was able to successfully **hear out** the low individual components or *partials* comprising a complex sound. The upper partials were more difficult to distinguish, but could be heard out more easily when attention was drawn to a target harmonic by alternately increasing and diminishing its intensity.

Based on these observations, von Helmholtz (op cit., p. 33) upheld the role of the ear as a Fourier analyzer and extended Ohm's law to read:

"Every motion of the air, then, which corresponds to a composite mass of musical tones is, according to Ohm's law, capable of being analysed into a sum of simple pendular vibrations, and to each such single simple vibration corresponds a simple tone, sensible to the ear, and having a pitch determined by the periodic time of the corresponding motion of the air".

2.2.3 Modes of listening

While the ear is indeed able to hear out some of the components of complex tones as observed by von Helmholtz and verified by Plomp (1964) and Plomp and Mimpen (1967), the normal, everyday mode of hearing is observed to be not so analytic.

We usually hear a complex sound produced by a source such as a

musical instrument as being a fused "image" (McAdams, 1984 b) or "entity" (Hartmann, 1988), with characteristic overall properties such as pitch, timbre and loudness. The fused entity is further, perceptually separable from other acoustic entities that may coexist in the environment with their own particular sets of components.

The ear is somehow able to fuse together the partials belonging to one complex sound, while separating them from those belonging to another sound. As stated by Plomp (1976, p.2), "apparently the ear is able to analyze a **compound** sound into **complex** rather than **simple** tones". It is able to take an input that is "a tumbled entanglement of the most different kinds of motion, complicated beyond conception" (von Helmholtz, op cit., p.26) and "decide which elements belong together, or come from the same source, and which elements come from different sources" (McAdams, 1984 b).

Von Helmholtz (1877/1954) addressed this apparent paradox of simultaneous segregation and grouping processes by suggesting that there are two different "grades" in our becoming conscious of a sensation, the "**synthetic**" and the "**analytic**":

"... the lower grade of this consciousness, is that where the influence of the sensation in question makes itself felt only in the conceptions of form of external things and processes, and assists in determining them. This can take place without our needing or indeed being able to ascertain to what particular part of our sensations we owe this or that relation of our perceptions. In this case ... the impression of

the sensation in question is perceived synthetically. The second and higher grade is when we immediately distinguish the sensation in question as an existing part of the sum of the sensations excited in us. We will say then that the sensation is perceived analytically." (p. 62)

Both simultaneous and sequential comparisons of acoustic features seem to take place in the sifting of spectral components into subgroups and their fusion within a subgroup into an entity. This grouping operation appears to hinge on the ability of the auditory system to listen both analytically and synthetically.

2.3 Frequency analysis

2.3.1 Limitations of Ohm's law

Plomp (1964) and Plomp and Mimpen (1967) probed the auditory limits of Ohm's law using an approach similar to that adopted by von Helmholtz, but with greater control over the stimulus variables. Their experiments aimed to determine how many components of a 12-component harmonic complex tone could be "heard out" separately as a function of fundamental frequency (F0) over the range 44 - 2000 Hz. For each F0, the percentage of correct responses diminished monotonically for increasing harmonic number n, i.e. the lower harmonics were easier to identify. As shown in **figure 2.1**, the average

number of harmonics audible across the range of F0 was 5-7. Plomp (1976) cautions however, that "there are substantial differences among subjects in their ability to identify partials" (p.4). This can be seen from the spread of data points in the figure.

Based on these data, Plomp and Mimpen calculated the minimal frequency separation required for correct identification of a partial. As shown in **figure 2.2**, at frequencies below 400 Hz, this limit was a constant value of 60 Hz. For higher frequencies, it was found to be a factor of about 15-20 %. The lower *frequency* components of a complex are thus better resolved by the auditory system, than are the higher components.

2.3.2 Critical concepts in hearing

The 15-20% frequency difference limit found by Plomp and Mimpen for resolution of harmonics of a complex agrees in magnitude with a **"critical band".** Fletcher (1940) proposed the concept of an auditory filter termed a **"critical band"** based on the results of studies of **auditory masking**, that also focus on the ability to analyze and discriminate the various components in a mixture of sounds.

The American Standards Association (1960) defines "masking" as the process by which the threshold of audibility of one sound is raised by the presence of another (as well as the amount (in dB) by which it is raised).

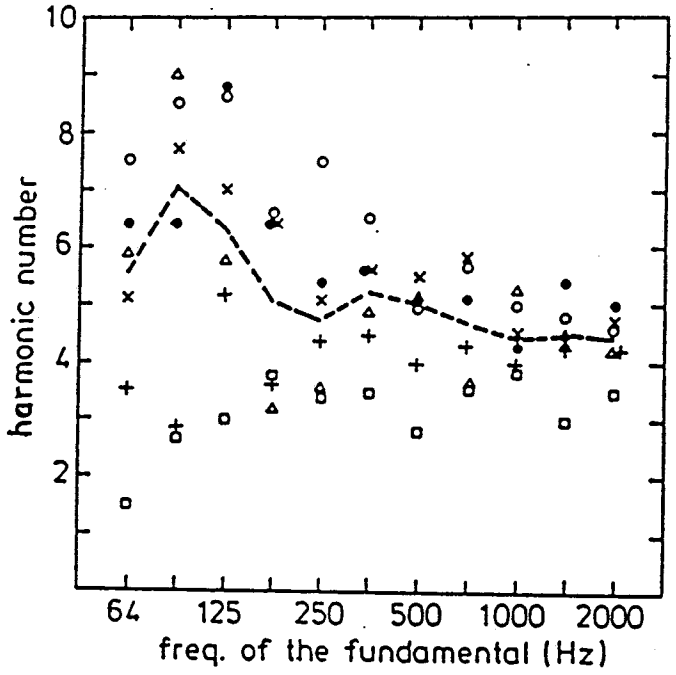


Figure 2.1 Number of harmonics of a 12-component complex tone identified by 6 subjects as a function of fundamental frequency. The dashed line represents the median. [From Plomp and Mimpen, 1968].

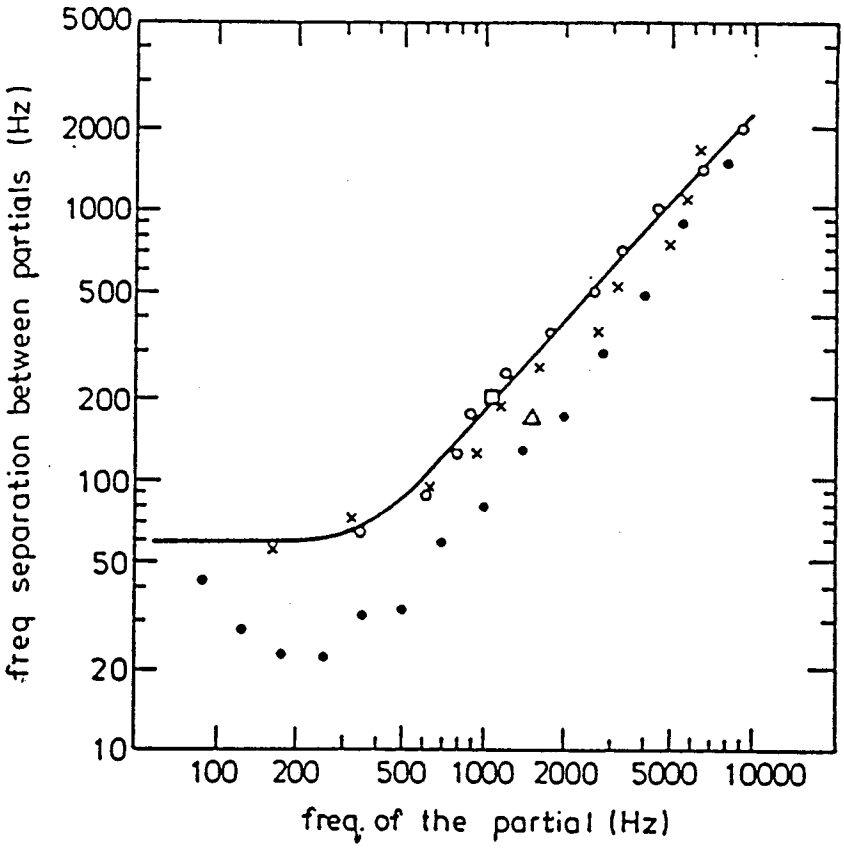


Figure 2.2 Data from studies by Plomp (1964) and Plomp and Mimpen (1968) showing the frequency separation between partials of a complex sound required for correct identification. The open circles (fitted with the curve) are for a 12-component harmonic complex (6 subjects), the crosses for an inharmonic complex (2 subjects). The solid points are for a 2-component complex (same 2 subjects). The open square and triangle represent the average data of a similar experiment by Soderquist (1970), with inharmonic tone complexes for 4 non- musicians and 4 musicians respectively. [Taken from Plomp, 1976].

Early studies of the masking phenomenon observed that frequency proximity was an important factor in the obscuring of one sound (the "signal") by another (the "masker"). This was revealed by deriving the "masking pattern" for a signal, which gives the dependence of the masked threshold for signal detection on the frequency relations between the signal and the masker. Frequencies of the masker in the vicinity of the signal were found to be more effective in masking, than those further removed. Fletcher suggested that the basilar membrane in the inner ear acts like a bank of bandpass filters with variable center frequencies (CF). In detecting a signal, the listener selectively attends to the filter whose CF is close to that of the signal. Only that part of the masker "noise" that is also passed through the filter is assumed to affect the threshold of audibility of the signal. The rest of the noise is not considered to have a deleterious effect on signal detection. The bandwidth of such a filter is referred to as the "critical bandwidth", or simply the "critical band" (CB).

The critical band has come to be accepted as **a basic characteristic of hearing.** It demarcates an area at the boundary of which listeners' responses to various features of complex sounds are observed to change. "Listeners react one way when the stimuli are wider than the critical band, and another way when the stimuli are narrower" (Scharf, 1970, p. 196).

The value of the critical bandwidth as a function of frequency is shown in **figure 2.3** (from Moore, 1982). The data are based on results

of several different measurements by different investigators. These data resemble the form of the function shown in **figure 2.2**, showing frequency separation (or bandwidth) to be a constant up to some mid-frequency value (400-1000 Hz) and then changing in proportion to the center frequency. Plomp and Mimpen thus concluded that "harmonics are distinguished only when their frequency separation exceeds critical bandwidth" (p. 767).

Frequency analysis, under the more general umbrella term "frequency selectivity", has been the focus of investigation in a vast number of studies. Some of these have concentrated on the physiological bases for frequency analysis and frequency coding in the auditory system. Others have adopted psychophysical approaches to measure aural frequency analysis and probe the limits of such resolution. A number of these studies are discussed in a review paper by Jesteadt and Norton (1985) and in a volume of papers edited by Moore and Patterson (1986).

2.4 Frequency synthesis

In contrast to the considerable research devoted to the issue of frequency analysis in hearing, the issue of frequency synthesis has received relatively little direct attention.

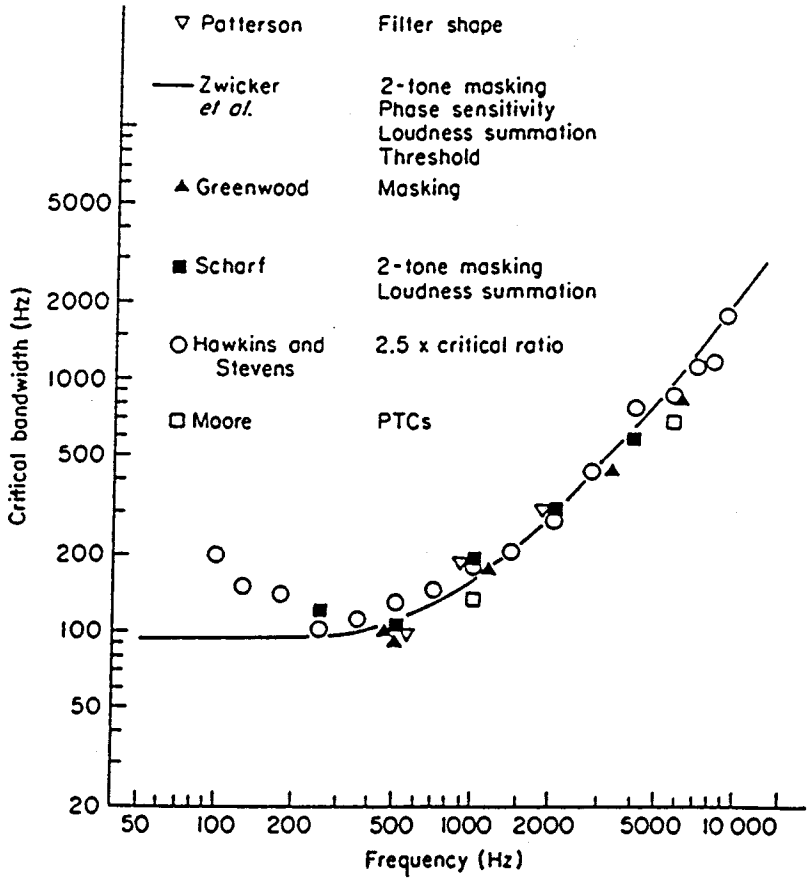


Figure 2.3 Value of the critical bandwidth as a function of frequency, compiled from different sets of measurements by Scharf (1970). [Taken from Moore, 1982].

The absence of frequency analysis is notable in two types of phenomena observed routinely: the mutual *cooperation* of spectral components to yield fused complex tones, and the mutual *interference* of components to yield diffused percepts such as beats, roughness and dissonance (Nordmark, 1978). These latter percepts directly represent the bounds of the frequency- analyzing ability of the ear and have been investigated to some extent in studies of modulation and the perception of phase relations (Goldstein, 1967; Mathes and Miller, 1947; Terhardt, 1968; Zwicker, 1952). The perception of complex tones on the other hand, represents a situation where the ear is *capable* of doing a frequency analysis, yet does not.

In some situations, both these aspects of component interaction may be manifested, as in complex sounds in which some components fuse together to yield a unified tonal percept, that may however be accompanied by fluctuations perceived as roughness or "warbling" or "fluttering" (McClelland and Brandt, 1969).

Theories of pitch perception have addressed the issue of synthetic listening to some extent, in describing how an array of frequency components is operated on to yield a unitary pitch for a complex sound as a whole (Goldstein, 1973; Terhardt, 1974). The larger question of what causes fusion of components to occur at all, was speculated on by von Helmholtz (1877/1954) in his book on "Sensations of tone" :

"... there are many circumstances which assist us first in separating the musical tones arising from different sources, and secondly, in keeping together the partial tones of each separate source. Thus, when one musical tone is heard for some time before being joined by the second, and then the second continues after the first has ceased, the separation in sound is facilitated by the succession of time. We have already heard the first musical tone by itself, and hence know immediately what we have to deduct from the compound effect for the effect of this first tone ..." (p. 59).

"... all these helps fail in the resolution of musical tones into their constituent partials. When a compound tone commences to sound, all its partial tones commence with the same comparative strength; when it swells, all of them generally swell uniformly; when it ceases, all cease simultaneously. Hence, no opportunity is generally given for hearing them separately and independently" (p. 60).

Until the last decade, very little empirical evidence existed to corroborate these insightful speculations made by von Helmholtz. This abyssmal situation has been ameliorated to some extent by the emergence of a new genre of researchers who have valiantly worked to shed light on the related phenomena of fusion of components of complex sounds and the segregation of such complexes from each other when presented simultaneously.

A study by Scheffers (1983) investigated cues that enabled

identification and discrimination of concurrently-presented vowels. Darwin (1984) has also investigated the factors enabling spectral components to fuse together and form a vowel and the constraints on such perception when other sounds are presented in addition to the vowel. McAdams (1984 a, b) took on the task of investigating specific acoustic determinants responsible for "spectral fusion" and "spectral parsing" of simultaneously-presented frequency components. Beerends (1989) studied the perception of pitches of simultaneous complex tones.

Some of these studies evolved from a more general research effort focussing on the practical problem of separating speech from interfering sounds (Duifhuis, Willems and Sluyter, 1982; Parsons, 1976). Others have evolved from a more epistemological effort to understand the processes and criteria involved in auditory perceptual organization in general (Bregman et al., 1983; Bregman and Doehring, 1984; Houtsma and Beerends, 1986).

Synthetic listening has been described here as the process by which a group of acoustic components are "fused" together into a unified perceptual sound event. However, this does not imply that such fusion is solely dependent on characteristics of concurrently- presented components. There are many indications that **contextual processing** of sequential information can influence the fusion of simultaneous components just as properties of fused sounds can influence perceptual grouping of sequences.

There are several interesting studies in which these two aspects of

auditory organization have been played off against each other (Bregman and Pinker, 1978; Dannenbring and Bregman, 1978; Deutsch, 1982). The recent volume on "auditory scene analysis" by Bregman (1990) reviews many such studies.

The following sections are devoted to describing **simultaneous** fusion. It is acknowledged that sequential context may play a role in aiding or hindering such fusion, but details of the interaction are not discussed at present. Chapter 6 returns to this issue of competition between simultaneous and sequential organization of sound.

2.4.1 Factors contributing to the perceptual fusion of complex sounds

The importance of similarity in the behavior of partials leading to their being construed as belonging to a composite musical sound as mentioned by von Helmholtz, was verified by McAdams (1984 a, b). He found that correlation of factors such as amplitude and frequency modulation across members of a spectral subgroup, contributes to their being perceptually fused into a unified "source image". The metaphor **auditory image** is used to describe "a psychological representation of a sound entity exhibiting an internal coherence in its acoustic behavior". The formation of a fused image simultaneously facilitates its perceptual separation from other spectral components that may be present in the auditory environment.

The cues listed below are viewed by McAdams (1984 a, p. 312; 1984 b, p. 40) as being the most "efficacious" in contributing to the related operations of formation and separation of source images. (Similar factors were listed by Moore (1982, p. 189) as cues for the separation of auditory "objects") :

1. (apparent) spatial location
2. harmonicity of spectral content
3. separation of pitches
4. coherence of frequency changes
5. coherence of amplitude changes
6. onset/offset asynchrony
7. stability and/or recognizability of spectral form (as conveyed by a complex coupling of amplitude and frequency modulation).

Of these parameters, harmonicity, coherence of low-frequency frequency modulation and the stability/recognizability of spectral form were directly studied in McAdams' experiments. Some of his observations, supplemented by those of other investigators, are reviewed below in the context of the cues listed above. It should be noted, however, that these cues do not comprise an exhaustive list of factors that affect fusion. Further, they need not be independent of each other, and may **cooperate or compete** in eliciting fusion or segregation

of components.

1. Apparent spatial location

A common perceived spatial location of spectral components can serve to imply to the listener that they are generated by the same sound source and should thus be fused together. Conversely, sounds emanating from different sources can be perceptually separated by the difference in location of their sources (Cherry, 1953).

In some contrived situations in which spectral components are distributed across ears (as in "dichotic" presentation), a conflict may arise in grouping by perceived spatial location and grouping based on some other criterion like harmonicity (Houtsma and Goldstein, 1972; Hall and Soderquist, 1975). For most natural sounds however, the common spatial location of a source is a useful cue in grouping the generated components into a unified percept.

2. Harmonicity of spectral components

Many natural complex sounds (e.g. the human voice) comprise "harmonic" spectral components that are multiples of a common fundamental frequency. These sounds are typically heard as being fused and characterized by a distinct pitch (Goldstein, 1973; Terhardt, 1974; Wightman, 1973). Inharmonic sounds on the other hand, in which partials are not multiples of a common F0, appear to be less fused and have ambiguous pitches (de Boer,1956; Cohen, 1980; Mathews and

Pierce, 1980; Schouten, 1940; Slaymaker, 1970). A number of theories have been proposed to explain why a harmonic series is perceptually fused to yield a sound with a unitary pitch. These are discussed in a later section on pitch perception.

Just as a harmonic series implies a single source with a distinctive pitch, multiple series of harmonics of different fundamentals may evoke multiple pitches and imply the presence of more than one source. This is observed, for example, in experiments with sounds in which one component is "mistuned" from its harmonic frequency while others maintain harmonicity (Hartmann, 1988; Martens, 1984; McAdams, 1984 a,b; Moore et al., 1985; Singh, 1989). The components that maintain harmonic relations are typically combined into one fused percept, while the mistuned component is heard as being segregated from the rest of the complex.

The harmonicity-seeking propensity of the auditory system has been exploited in the development of harmonicity-detector models that solve practical problems in pitch estimation and separation of mixed signals (Duifhuis, Willems and Sluyter, 1982; Parsons, 1976).

3. Pitch separation

Pitch separation (or more accurately speaking, a **difference in F0**) is an aid in the "sifting" of concurrently-presented vowel sounds. Scheffers (1983) found that accuracy of vowel identification was sensitive to changes in F0 from 0-6% (i.e., up to a semitone). The change

in pitch resulted in improved vowel identification. Larger changes in F0 (beyond 2 semitones), however, did not yield any further improvement. Brokx and Noteboom (1982) also reported similar improvement when F0 was changed up to 3 semitones (18% change). Whether these improvements resulted due to actual perceived differences in overall *pitch* per se, or because of perceived differences in pitches of individual components or in timbre, however, is a matter of some debate.

When the F0 of a harmonic complex such as a vowel is changed, the harmonics correspondingly change in their absolute frequencies, by the same proportion as the fundamental component. For different amounts of F0 change therefore, there will be different degrees of **spectral overlap** between the components of sounds presented simultaneously. Scheffers thus acknowledged that listeners could also compare changes in the *components* of the vowels or in the overall spectral "form", rather than the pitch per se, as cues for identification and discrimination of vowel sounds.

Influence of spectral overlap was also reported by Houtsma and Canning (1984) for a musical-interval identification task where the two complex tones comprising the interval were presented simultaneously. While perfect harmonic coincidence of components of the two tones did not impede judgment, overlap of the general spectral region of components made the task of separating them into the appropriate groups corresponding to the original tones more difficult. Houtsma and Beerends (1986) attributed this type of degradation in performance to

peripheral interference and lack of resolution of partials within the cochlea.

Chalikia and Bregman (1989) further studied the effect of F0 on perceptual segregation of simultaneous signals using "gliding" sounds in which F0 changed dynamically over time (as in *glisssandi*). Both, increased differences in F0, and F0 movement in time led to better separation. Further, "crossing" glides (in which the frequencies of components of the two sounds glided in opposite directions) led to even more significant improvement in identification of vowels presented simultaneously. This result was attributed to the commonality of frequency modulation among the gliding harmonics of one vowel, and the lack of such coherence across the harmonics of both vowels.

4. Coherence of frequency changes

It is a well-known strategy in music composition that parallel pitch movement will lead to a better "blend" of voices, while lack of parallel movement will enhance their separability. This has been verified empirically by Chalikia and Bregman (1989) as described above. In addition to the smooth, gradual change in frequency provided in "glide" stimuli of the type used by them, there is also evidence that correlations in the fluctuant, up-and-down deviations in frequency provided by typical "frequency modulation" (FM) can cue the separation and fusion of components.

An experiment by Rasch (1978) showed, for example, that FM

imposed on the components of one tone of a pair of complex tones presented simultaneously, improved its detection considerably (as much as 17 dB for a 5 Hz modulation rate).

The fact that many natural sounds exhibit "vibrato" or modulation in the fundamental frequency or frequencies of components has been noted by many observers (Beauchamp, 1974; Benade, 1976; Bjorkland, 1961; Fletcher et al. 1965; Fletcher and Sanders, 1967; Grey and Moorer, 1977; Lieberman, 1961; McAdams, 1984 a, b; Saldanha and Corso, 1964). Despite these deviations in frequency, a fused pitch is usually perceived for these sounds (Shonle and Horan, 1980). Since they do not seem to hinder fusion, the ecologically sound explanation for the existence of the deviations may be that they in fact **aid fusion** (Benade, 1986).

McAdams (1984 a, b) explored this possibility in experiments in which listeners were required to report differences in the perceived number of sources on comparison of sounds in which either **coherent or incoherent FM** was imposed on one or more partials. Under coherent modulation, all partials exhibited correlated "frequency behavior", as was the case when they were modulated by the same modulator. Such sounds remained fused perceptually, while those in which subsets of components were modulated incoherently (uncorrelated with the other components) appeared to segregate into more than one source.

"Coherence" of modulation *per se,* however, may not be a sufficient cue to imply fusion. It may be the interaction of FM with

another cue - such as harmonicity, for example, that promotes fusion. This was shown in other experiments by McAdams, in which the frequency of spectral components was modulated, using 2 *types* of FM. One type of modulation preserved the ***relative*** frequency relations between components, with all harmonics modulated to an extent proportional to the harmonic frequencies. The harmonicity of the complex was thus preserved. Another type of FM preserved the spectral spacing between components. In this case, all components were modulated to the same ***absolute*** extent, and the complex thus became inharmonic.

For ratio-preserving FM, the stimuli retained their perceptually fused character and were heard for the most part as timbrally rich tones with varying pitch. For FM that preserved component spacing but led to inharmonicity, the unitary image of the stimulus was lost and multiple sources perceived.

Bregman and Doehring (1984) also reported decreased fusion for gliding sounds in which the changing frequencies of components did not maintain harmonic relations with each other. Despite the "coherent" modulation of components gliding in parallel with respect to each other, fusion was undermined by the lack of harmonicity. A conflict of fusion and segregation cues was thus observed.

5. Coherence of amplitude changes

The perceptual fusion brought about by correlated changes in

frequency as described above, is an example of ***cross-channel coherence detection*** (where "channel" refers to the frequency- limited output of a critical-band-type filter). Such a mechanism allows comparisons and correlations of temporal behavior to be made across auditory channels (McAdams, 1984 b, p.94).

As was the case for dynamic changes in frequency, correlations in changes of amplitude over time can also bring about perceptual fusion of the components of a complex sound (Moore, 1982, p. 192-193).

McAdams (1984 b, p.51) cites an experiment by von Bekesy (1963) that showed that two pure tones presented separately to the two ears could be made to fuse perceptually by imposing a common amplitude modulation (AM) on them. The fusing effect of such comodulation has received considerable attention lately in studies of **"comodulation masking release"** or "CMR" (Hall et al., 1984).

In most studies of auditory masking, a consistent result that emerges is that the most important factor underlying signal detection is the signal-to-noise ratio at the output of the critical band centered at the signal frequency. Masker energy at regions remote from the signal frequency has traditionally been thought to contribute little to the masking of the signal (Fletcher, 1940). But recent experiments by Hall et al. (1984, 1986, 1987), McFadden (1986), Green (1988 a) and others, have shown that under certain conditions, the auditory system is capable of adopting a wideband-analysis strategy in signal detection. This is observed, for example, when modulated noise is used, or where there is

across-frequency coherence of amplitude as manifested by a common temporal envelope (Buus, 1985; Hall, 1987). The across-frequency coherence of the waveform envelope is found to be beneficial in rendering the signal detectable. The comodulated noise bands are grouped together as one source, while the signal is segregated as a separate perceptual entity and thereby "released" from masking. Hall et al. (1984) thus suggest that "the auditory system uses across-frequency analysis of temporal modulation patterns to help register and differentiate between acoustical sources" (p. 56).

Spectral integration based on common amplitude modulation has also been reported by Bregman et al. (1985). They report that two complex tones generated via AM were judged to be more fused when they shared the same modulation frequency, even if the spectral components thus generated were not harmonically related to each other. Apparently, the common periodicity served to fuse the complexes.

This type of fusion of inharmonic partials by compensating amplitude relations was also demonstrated by Cohen (1980). She was able to bring about perceptual fusion of inharmonic partials by imposition of a synchronous, exponential amplitude envelope on them.

6. Onset/offset asynchrony

The importance of synchrony in contributing to the perceptual fusion of partials of a complex tone was alluded to by von Helmholtz as cited above (section 2.4). The use of *asynchrony* to differentiate voices

in music is also well known. Three contemporaneous experiments by Bregman and Pinker (1978), Dannenbring and Bregman (1978), and Rasch (1978), provided empirical evidence for the importance of synchrony in facilitating fusion and of asynchrony in facilitating segregation or *fission.*

The experiment of Bregman and Pinker (1978) showed that two pure tones are more likely to be perceptually fused into a complex tone with a "rich" timbre when they are temporally synchronous. Dannenbring and Bregman (1978) explored details of this relation between asynchrony and fusion further. They showed that asynchronies in onset and offset of components led to segregation only when the deviant component *led* or *lagged* behind the other components (i.e. asynchrony in onset or offset that occured *during* the time period of occurrence of the other components did not facilitate segregation of the temporally deviant component). Lead asynchrony also yielded fission more readily than lag asynchrony.

Rasch (1978) tried to quantify the advantages of asynchrony in hearing out simultaneous notes (as in polyphonic music) using a masking paradigm. When two complex tones were presented overlapped in time, the detection of the note with the higher F0 was facilitated by as much as 50 dB when their onsets were made to differ by 10-30 msec, versus the case where they were exactly synchronous.

The time-variant spectra of most natural instruments show some differences in features of the temporal amplitude envelopes of

components, such as asynchronies and differences in rates of rise of amplitude (Grey, 1975; Grey and Moorer, 1977). Despite these variations, a fused tonal percept is typically evoked. McAdams (1984 b, p. 50) explains this seemingly-paradoxical situation, reasoning that the observed asynchronies are usually fairly small (≤ 30 msec), and further, are camouflaged by the noise typically generated during the "transient", initiation stage of sound production. While not hindering fusion, these subtle differences amongst partials are important, however, in conveying the characteristic quality or timbre of the sound (Berger, 1964; Saldanha and Corso, 1964).

7. Stability/recognizability of spectral form

The recognition of spectral form is an operation carried out routinely by the auditory system in the processing of vowels, and in the identification of instrumental timbres.

The spectral envelope of a sound outlines the pattern of peaks and dips showing enhanced or reduced amplitudes at different frequencies. Such peaks, termed "formants" by Hermann (1890; cited in Winckel, 1967), correspond to regions of resonance in the sound transmission system.

Vowel sounds are typically characterized by a specific formant structure correlated with resonances in the vocal tract. These resonance frequencies are considered to be the *major* information- bearing elements in the assignment of phonemic identities (Carlson, Fant and

Granstrom, 1975; Miller, 1984).

Recognition of resonance structure can also serve as a cue for the perceptual grouping of components. Resonance structure can be inferred by repeated encounters with sounds produced via a particular source-filter configuration. Spectral components may then be fused together if they appear to follow the inferred spectral form (manifested by the shape and position of the spectral envelope).

As hypothesized by Bregman (1990, p. 575):

"if changes in the intensities of certain harmonics accompanied changes in the fundamental in a way that was consistent with their being passed through the same resonance system, this might contribute to their being grouped together".

The importance of frequency and amplitude modulation in promoting fusion of spectral components was supported by McAdams (1984 a,b), who showed that a coupling of the spectral envelope with these two types of modulations can serve to define a spectral group that conveys features such as phonemic identities. "Coupling" here implies that the amplitudes of the components are constrained to follow the shape of the spectral envelope, when they are modulated in frequency. The "tracing" of the envelope thus provided, may aid the identification of the spectral group as comprising a perceptual unit such as a vowel sound.

Figure 2.4 (taken from McAdams 1984 a) illustrates how the amplitudes of spectral components follow the contour of the formant structure when vibrato is introduced in a vowel sound.

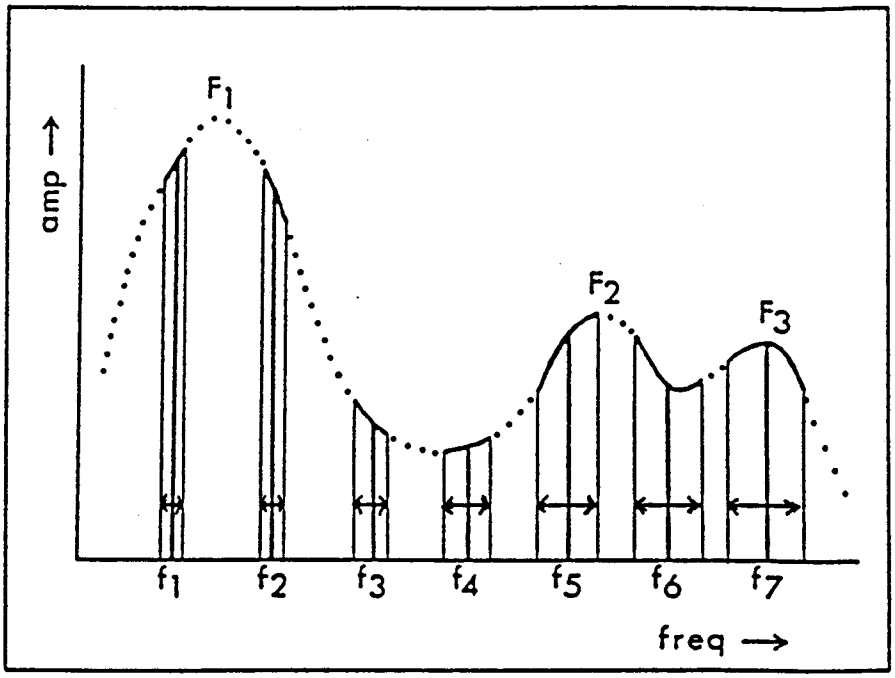


Figure 2.4 Constrained modulation of harmonics of a 3-formant resonance structure enables correct inference of the spectral form (represented by the dotted line). The solid portions of the line indicate the amplitude trajectories of components as they are modulated in frequency. The tracing out of portions of the spectral envelope "point" toward formant peaks. [From McAdams, 1984 a, b].

This "pointing" of the modulated components can be of assistance in enabling the correct inference of the formants. For sounds with high values of F0, the wider spacing (and thus reduced density) of components can lead to misjudgment of the spectral form. Tracing of the envelope by the introduction of modulation is particularly helpful in defining the spectral envelope for such sounds. **Figure 2.5** (also from McAdams, 1984 a) shows how ambiguity in the spectral form of the vowel /a/ at a high F0 may be dispelled by modulating the components.

In addition to dynamic features such as modulation of frequency and amplitude, the "naturalness" of more static aspects of spectral form may also affect perceived grouping of components. Such naturalness may be manifested in terms of the realistic *plausibility* of the implied resonance structure. Thus, an abrupt defiance of an otherwise smooth spectrum by sharp peaks may lead to the rejection of a component as not "belonging" to the spectral group.

A group of spectral components that exhibits a falling off in intensity for higher frequencies is also more likely to be fused to imply a single source image, than a spectrum that does not have such a "rolloff" of intensity. Benade (1981) claimed that "good", "well carrying" instruments are typically characterized by a spectral rolloff in the amplitude of components of the order of $1/f^3$ (f=frequency of component). Such a rolloff of high harmonics serves to reduce inter-component masking in upper critical bands while also reducing "roughness" and facilitating the tracking of pitch and timbre.

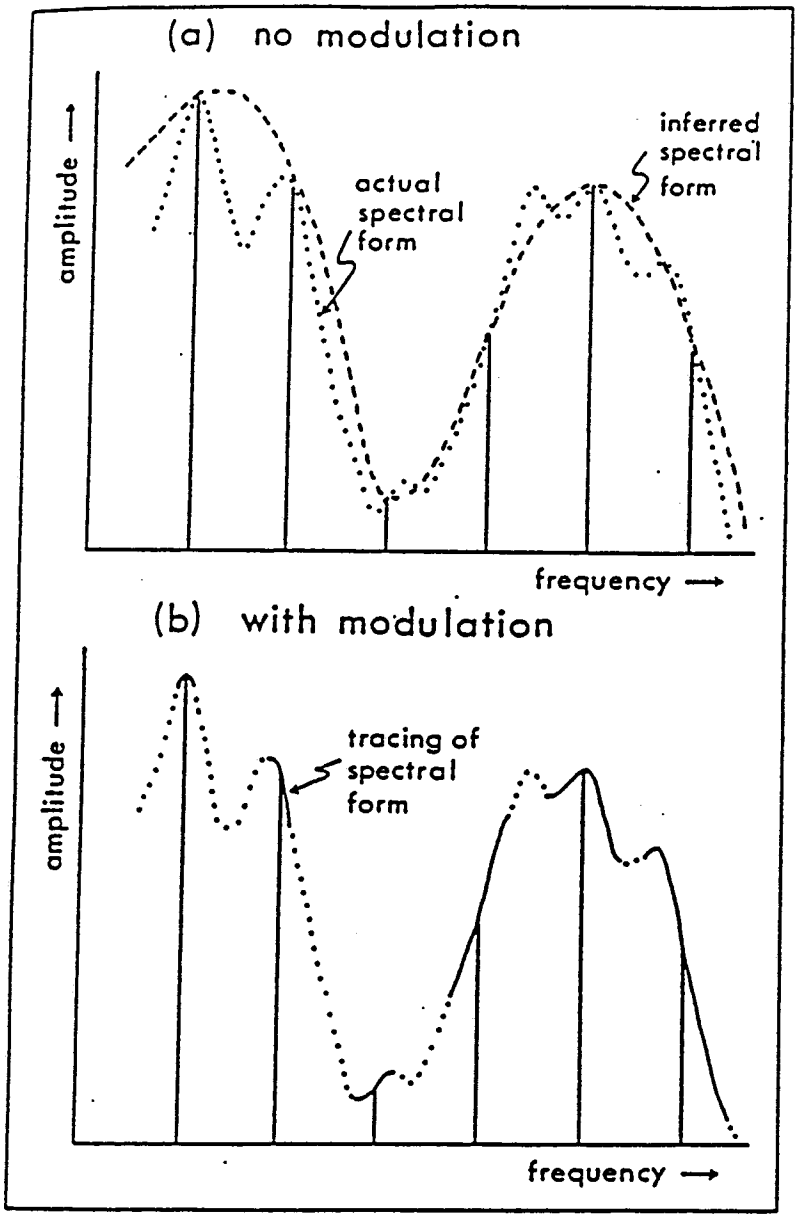


Figure 2.5 Spectrum of the vowel /a/ plotted with few harmonics of a high fundamental frequency. Without modulation, (a) shows that inferred spectral form may be quite different from the actual spectral form. With modulation, a clearer indication of the spectral form is available via the amplitude trajectories of the modulated components as shown in (b). [From McAdams, 1984, a,b].

2.4.2 Interaction and tradeoffs between fusion cues

The spatial, spectral and temporal cues listed above can compete or collude in evoking perceptual fusion. Some examples of this conflict or symbiosis have already been given above. A few others are described below:

i) Interaction of FM and spatial cues

A striking audio example provided by McAdams (1984 b; created by Reynolds and Lancino) manifested a situation where spatial separation of sounds was superseded by fusion-inducing FM. The odd and even harmonics of a sampled oboe sound were fed through separate channels to different loudspeakers. Despite the spatial separation of components, the coherence in the FM initially imposed on all the harmonics led to a fused image of an oboe, localized between the two speakers. However, when the odd and even subsets of partials were modulated *incoherently* with respect to each other, the original sound split into two "voices" : a clarinet-like tone localized in one speaker, and a soprano-like voice an octave above it, localized at the other speaker. The common FM within a subset served to keep it fused, while the uncorrelated modulation across subsets, coupled with the spatial separation, served to segregate them perceptually.

Odd harmonics are characteristically dominant in the spectra of clarinets (Benade, 1976). Their fusion thus resulted in a clarinet-like

timbre. The fusion of even harmonics on the other hand, implied a new harmonic series with all components at double the frequencies of the original series. A new voice thus emerged at the octave above the original pitch.

ii) Contralateral fusion and assignment of phonetic identity

An interesting dichotic phenomenon reported by Broadbent and Ladefoged (1957) demonstrated a tradeoff between spatial location and spectral form in the fusion of spectral components to convey a consonant-vowel syllable (/da/). The first two formants (F1 and F2, and transitions) of the CV syllable were presented separately to the two ears. When presented in isolation, the input in each ear was heard as being a non-speech sound and was localized accurately. However, with simultaneous presentation, the inputs at the two ears fused together to establish a unitary percept identified as a /da/, with loss of accuracy in localization of the appropriate components to the correct ear. This phenomenon was studied further by Rand (1974) and Cutting (1976). The latter, manipulated factors such as synchrony relations and F0 of the material in the two ears. Despite temporal offsets of up to 20 msec and F0 shifts of $\leq$ 80 Hz, the phonetic identity established by the dichotic stimulus was maintained. However, the perception of a unitary source was sometimes replaced by one of multiple sources present simultaneously. This paradoxical situation of concurrent fusion and

fission arises because of the conflicts between different cues: The recognizability of spectral form allows a unified phonetic identity to be established across ears. The difference in F0 and spatial location of the individual components on the other hand offers a competing cue that suggests the presence of more than one source.

2.4.3 Conclusion

The perceptual fusion of simultaneous components thus appears to depend on the relations between primary dimensions of sounds, such as frequency and intensity and their interaction with temporal and spatial factors. Correlated changes along different dimensions over time typically indicate a common source origin that leads to components being grouped together into a unified percept. However, the cues responsible for bringing about fusion or separation interact in a **context-dependent** manner. Conflict or cooperation of cues can serve to inhibit or facilitate the fusion of individual components into unified "events" with higher-order perceptual attributes associated with the fused ensemble as a unit. Some of these attributes are discussed in the following sections.

2.5 Perceptual attributes of auditory events

The perceptual fusion of a number of components into an "entity"

usually results in the loss of the individual identities of the components, in the service of an emergent perceptual property of the complex as a whole (Cutting, 1976). The principal perceptual attributes typically associated with unitary events are: **pitch, loudness, apparent duration, and sound quality or *timbre*** (Hirsh,1988, p. 379).

While loudness and duration are more transitory aspects of sound, vulnerable to being obscured by environmental conditions, pitch and timbre are more resilient, distinctive perceptual features. The timbral identity of a sound source is retained despite a variety of disturbances in the listening or transmission path (Wente, 1935; Risset and Wessel, 1982). In fact, under normal listening conditions (as opposed to listening in the laboratory), a signal can undergo several kinds of modifications and degradation, before a listener perceives a marked change in timbre (Charbonneau, 1981). Pitch, too, is a survivor. The pitch of a signal can be perceived as unchanged despite variations in listening conditions and in details of the stimulus (Schouten, 1939; Rasch and Plomp, 1982).

Pitch has enjoyed a particularly hallowed status in music as the traditional bearer of structural information. **Timbre** on the other hand, has traditionally been relegated to an embellishing, ornamental role. Only lately has it been put to use as another structural element capable of conveying musical schemes (such as the "klangfarbenmelodie" of Schoenberg, the sound masses of Webern and Ligeti, and the "minimalist" compositions of composers like Reich and Glass).

Pitch perception has also been a major subject of investigation in

psychoacoustics. Pitch is an auditory percept that lends itself well to quantifiable comparisons. Pitch can be differentiated along both ordinal scales (from low to high), and ratio scales (defining intervals such as the musical octave, fifth etc). Loudness and perceived duration, too, can be ordered along some type of scale of magnitude (e.g. soft to loud, short to long). Timbre differences, on the other hand, are difficult to quantify and are typically described by nominal, semantic labels characterizing the sound source (e.g. names of instruments; voice qualities such as "sharp", "bright", "dull"; or source labels such as "tinny", "hollow", "metallic" etc). Consequently, the perception of pitch has received a greater impetus in research studies, than has the perception of timbre. The primary reason for this disparity has been a limited understanding of the physical determinants of timbre as well as the lack of a common perceptual reference with respect to which timbre differences can be described.

Some of the prior research on pitch and timbre perception is reviewed briefly in the next two sections. Further coverage of these and other perceptual attributes of sounds and additional references can be found in publications by Plomp (1976), de Boer (1976), Deutsch (1982), Moore (1982), Green (1988 b), Hirsh (1988) and Handel (1989).

2.6 Pitch and timbre: Parallel lines of investigation since von Helmholtz

At the very outset of his voluminous tome on "Sensations of tone", von Helmholtz (1877/1954, p.10) stated that the tones of music could be distinguished by three features:

1. by their force,
2. by their pitch,
3. by their quality.

(In present-day terminology, "force" would correspond to perceived loudness, while "quality" would correspond to perceived timbre).

In assigning physical determinants to these musical features, von Helmholtz looked both toward the acoustic waveform and the Fourier spectrum for cues:

From a **waveform** point of view, von Helmholtz observed that the loudness (or "force") of sounds seemed to be correlated with the amplitude of vibration, while the pitch appeared to be correlated with the rapidity of vibration (i.e. "on the number of vibrations completed in a given time", or equivalently, by the inverse of the "period" of vibration). "Quality" was considered to be a more enigmatic percept that von Helmholtz explained by a process of elimination: Since force was known to depend on the amplitude of vibration, and pitch depended on the rapidity of vibration, there was "nothing left to distinguish quality but vibrational form" (p.65).

From a **spectrum** point of view, von Helmholtz observed that the partials of sounds produced by most instruments were multiples of a strong, dominant fundamental component. The frequency of this lowest

partial was also exhibited in the periodicity of the compound waveform. Von Helmholtz thus equated the unitary pitch of the complex with the pitch of the lowest partial, while cautioning that:

"a note has, properly speaking, no single pitch, as it is made up of various partials each of which has its own pitch" (p. 24), and that "the sensation of a musical tone is compounded out of the sensations of several simple tones" (p. 56).

Since the lowest partial was believed to be the "fundamental" component, the frequency assigned to represent the "synthetic" pitch of the whole complex was taken to be equivalent to that of the fundamental (usually denoted by "F0" or "fo").

With the lowest partial held accountable for the pitch of a complex tone, von Helmholtz hypothesized that "the form of the vibration of the air" that led to the perception of **"quality"** of musical tones must depend on the distribution of upper partials or overtones (op cit., p. 65). A corollary of this assertion was the hypothesis that "different modes of combining the upper partial tones correspond to characteristic varieties of musical quality" (op cit., p. 69).

As stated at the outset of this chapter, the distribution of partials and their relative intensities constitutes the "spectrum" of a sound that can be represented in the time domain by the function

$$f(t) = \sum a_n \sin(2\pi n f_0 t + \phi_n)$$

comprising components "n" of frequency ($=nf_0$), amplitude ($=a_n$) and

phase (ϕ_n).

Having established the relation between pitch and f_0, and timbre (or quality) and the amplitude spectrum (characterized by the distribution of a_n), von Helmholtz went on to investigate the influence of the phase relations (ϕ_n) on the evoked pitch and timbre.

Based on personal observations with equipment comprising a set of tunable resonators, von Helmholtz concluded that neither pitch, nor the quality of complex tones depended upon the phase differences between partials. He even went so far as to "lay down the important law that *differences in musical quality of tone depend solely on the presence and strength of partial tones, and in no respect on the differences in phase under which these partial tones enter into composition*" (op cit. p. 127).

The amplitude spectrum thus reigned as the major conveyor of timbre and the fundamental component as the conveyor of pitch. A simple explanantion ensued to describe how sounds of different tonal quality could have the same pitch: Two sounds could have the same pitch but differ in timbre because of the difference in overtone structure as manifested in the spectrum. Or two sounds could have a similar timbre but differ in pitch because of a different value of the fundamental frequency. Timbre and pitch were therefore considered to be **separable, independent aspects of tone sensation.**

The perceptual and conceptual separation of pitch from timbre led to two parallel lines of investigation in the century that followed.

The apparently straightforward relation between pitch and frequency, and the ease with which pitch could be compared and judged by listeners, led to its being used to probe properties of the hearing mechanism. Timbre on the other hand was more difficult to describe, both physically and perceptually.

The complex physical behavior of musical instruments and the ensuing spectra of the sounds produced seemed to be highly correlated with the timbre perceived. Several physicists followed the lead of von Helmholtz and tried to approach the problem of timbre specification by studying characteristics of musical instruments and the sounds they produced. In comparison with the study of pitch, the study of timbre thus initially took on a more physical, rather than psychological or physiological orientation.

2.7 Timbre: Physical and psychophysical investigations

2.7.1 Search for the physical correlates of timbre

Musical instruments typically retain their separate perceptual identities even when played in reverberant rooms, separately or in an orchestral context, live or via recording (Eagleson and Eagleson, 1947; Benade, 1983). This observation raises the question of "what" exactly it is that imparts the perceptual sturdiness to an instrumental timbre?

The "classical" view that tone quality or **timbre was derived**

from the spectrum was propagated well into the present century. The characteristic spectra obtained for different musical tones were considered to arise because of the relative strengthening or attenuation of partials dependent on the resonances of musical instruments. Given the success of formants as descriptors of vowel identity in speech, Fletcher (1934) invoked the idea of using "formants" as descriptors of musical quality. Spectral analyses of sounds of different instruments were thus scrutinized in search of salient features of the underlying formant structure that could be responsible for the perceptual invariance of timbre across different contexts (Benade, 1976, 1983, 1986; Culver, 1956; Miller, 1926; Olson, 1967).

Most of these analyses, however, came up with frustratingly inconclusive data that failed to capture the "essence" of the spectral features that were presumed to contribute to the perceptual sturdiness of timbre. The analyzed spectra did not appear to show any consistency that could be noted as contributing to the timbre, even when measurements were obtained from sounds produced by the same instrument.

Saunders (1946) analyzed different wind instruments in search of consistent "formants" or resonances in the spectrum that would be indicated by dominance of different harmonics in the spectrum. He lamented that "a search was made for formants in all the instruments" (clarinet, oboe, flute, english horn, french horn), ... "without success".

The elusiveness of formants in analyses of musical instruments

prompted Bolt (1948, p.66) to write a letter entitled "Wanted - The Formant - dead or alive" to the Journal of the Acoustical Society of America. In this colorfully-worded letter, Bolt commented on the "unquestioned" alliance of formants and vowels, and the "possible bigamous relations with musical instruments" !

The inconsistency of spectra obtained from analyses of instruments led Culver (1956; cited by Schmid, 1977) to wonder aloud why it was that four different notes produced on the same violin should evoke the same timbre when the harmonic structure obtained by spectral measurements was very different. The spectral representations apparently failed to capture the common factor that led to the invariance of timbre.

Benade (1983) summed up the frustration of these analytic approaches with the statement "instrumental spectra are very variable indeed, yet the timbre associated with the instrument appears to be perceived as invariant". He also pointed out that spectral measurements were subject to considerable variability depending on situational factors such as the relative positions of the measuring device and the instrument being tested and the time span of measurement. He proposed that a "room-averaged" spectrum derived from the combined spectra of an instrument playing different notes at different locations would be better representative of the spectral characteristics of the instrument.

The inadequacy of early spectral analyses as representations of instrument timbre was demonstrated via an **analysis-synthesis**

approach used by Risset and Mathews (1969). The rationale of this approach is that an accurate analysis of the physical aspects of timbre should enable a resynthesis of a sound that renders it indistinguishable from the original analyzed sound. If the resynthesized version of the sound (assuming accurate synthesis techniques) is perceived as being timbrally different from the original, then the analysis failed to retain the crucial aspects of the timbre-bearing features of the sound (Risset, 1978).

One reason for the failure of the older analyses in passing this analysis-synthesis test was the lack of information about the **dynamic** aspects of sound variation (Risset and Wessel, 1982). As mentioned before, the analyzing equipment and methods for obtaining the frequency spectra in earlier analyses entailed ignoring or averaging out the temporal characteristics of amplitude evolution of partials and modulations in frequency and amplitude over the course of the sound. These dynamic, "transient" features of sounds have since been recognized as playing a major role in the identification of instruments (Berger, 1964; Clark et al., 1964; Grey, 1975; Luce, 1963; Saldanha and Corso, 1964; Wessel, 1979).

2.7.2 Temporal features and timbre identification

Von Helmholtz had acknowledged that many of the "peculiarities" contributing to the quality of a sound depended on how the sound began and ended. He mentioned that the methods of "attacking and releasing"

tones can be particularly characteristic of a sound source. Beginnings and endings are "transient" features of sounds. The spectrum fluctuates wildly during the initiation of a sound. Further, different moments of sound initiation may differ considerably depending on the energy of excitation, the skill of the initiator etc. Due to the ephemeral nature of these dynamic aspects of sounds, von Helmholtz deliberately limited his investigation of timbre to studies of the "**steady-state**" spectrum of a sound that defines "the peculiarities of the musical tone which continues uniformly" (von Helmholtz, op. cit. p. 67).

Impairment in recognition of instruments brought about when initial segments of notes were removed was noted by Stumpf (1926, cited by Risset and Wessel, 1982). The role of "transients" in contributing to tone quality was also acknowledged by Young (1960). To test this idea empirically, Saldanha and Corso (1964) undertook a study designed "to evaluate the relative importance of transients, harmonic structure and vibrato as timbre cues in the absolute judgment of musical tones". "Absolute judgment" here refers to the ability to correctly identify a presented sound with the timbre of the instrument that produced it.

The stimuli presented comprised segments of recordings of various instruments. The original recorded samples were modified via tape-splicing techniques to give 5 types of stimuli comprising combinations of three portions of the sounds: the initial transients, the steady state, and the final transients. They verified that the presentation of the initial attack transients led to greater accuracy of identification,

and further reported the surprising result that the initial transients presented alone, resulted in *better* identification performance than presentation of the original, "complete" sound. The temporal features of sound initiation thus do seem to be crucially important cues contributing to the correct identification of the timbral "signature" of a sound.

2.7.3 Relative importance of spectral and temporal characteristics in identification of instrument timbre

The relative roles played by spectral and temporal features of a sound in enabling its correct recognition were investigated by Strong and Clark (1967). They first obtained identification judgments for synthetic spectra that were derived from analyses of real instruments. Then the spectral and temporal envelopes were **exchanged** between different "instruments" and changes in identification performance were observed. It was found that the spectral envelope was critical for the identification of instruments that are characterized by unique spectral envelopes (such as oboe, clarinet, bassoon, trumpet annd tuba). The temporal envelope became the more dominant cue for identification of instruments where the spectral envelope was not unique (as for the flute, trombone and horn).

This type of combination of properties of different instruments termed "cross synthesis" (Risset and Wessel, 1982), lends itself well to

the concept of a sound producer being represented by the combined contributions of a "source" and "filter" set (Huggins, 1952; Slawson, 1985).

The "source" is the seat of the excitation energy which is transformed in various ways, depending on characteristics of the "filter" response. The temporal aspects of a sound are considered to be related primarily to the excitation source while the spectral features represent the total combination of the excitation signal modified by the filter set. The auditory system appears to be well-equipped to separate the spectral and temporal aspects of sounds (Huggins,1952).

The source-filter model has been widely used in speech perception research (Fant, 1960) and in the development of speech analysis and synthesis techniques (Makhoul, 1975). The source-filter model and "linear prediction" algorithms developed for speech research hold great potential in music applications, particularly in the description of "tone color" (Moorer, 1978; Slawson, 1985) .

2.7.4 Auditory spectral filtering and phase perception: Timbral consequences

The dependence of timbre on the amplitude spectrum has always been well established. The role of phase on the other hand, has usually been underestimated. Von Helmholtz' observation that timbre did not appear to change despite large variations in the waveform

brought about by changes in phase, led him to state that phase differences between partials are unimportant in the perception of pitch and timbre (see section 2.6, p. 36).

Plomp (1970) however, tried to tone down the vehemence of the strong stance on insensitivity to phase usually attributed to von Helmholtz. Citing other observations recorded by von Helmholtz (1877/1954, p. 119), Plomp asserts that von Helmholtz "did not consider it to be impossible that, since harmonics beyond the 6th to 8th give rise to **dissonance and roughness**, a phase effect does exist for these higher harmonics" (*sic*. Plomp, 1970, p.400).

The influence of higher harmonics on the perception of roughness had indeed been verified by Lichte (1941). Perceived roughness was found to increase with the presence of *consecutive,* high harmonics beyond n=6.

"Roughness" has continued to be a popular term, used often in descriptions of sound quality. It appears to be related to the waveform, and thus may be induced by many factors that change the waveform, such as the location and amplitude relations beween components, as well as their relative **phase**. Temporal interference effects caused by limited spectral resolution appear to dominate the perception of roughness.

Both Licklider (1957) and Schroeder (1959), reported discriminable changes in timbre correlated with changed phases of harmonics. In keeping with earlier observations, Licklider reported that changes in the phase of lower components of a 16-harmonic complex had

less of an effect on discriminability, than did changes in phase of higher harmonics. The effect of phase was also more pronounced at low fundamental frequencies.

Both these observations are indicative of interference between components. High harmonics occupy higher frequency regions where the critical bandwidth is wide enough to include more than one harmonic. For low F0s, the spacing between spectral components would be small. The reduced spacing, and therefore increased spectral density would cause components to interact within critical bands in higher frequency regions, leading to changes in the temporal fine structure of the stimulus waveform.

The sensation of roughness appears to depend primarily on the relative fluctuations of the temporal amplitude envelope brought about when the frequency separation between spectral components is not too large (≤ critical bandwidth) (Mathes and Miller, 1947; Zwicker, 1952).

Mathes and Miller (1947) studied the "differences in sensation ... produced by changes in the phase alone of but a single or group of components". For a change in the phase of a single component ($\pi/2$ shift in the carrier frequency) of a 3-component complex obtained via AM, clearly audible changes in *roughness* were perceived. The change in waveform was described as "quasi-frequency modulation" (QFM) (after Stevens and Davis, 1938). Sensations of smoothness and roughness due to simultaneous phase changes in *multiple* components appeared, however, to be combined in ways that tended to mask each other.

Phase effects have also been reported by other investigators, although not always in the context of timbre perception (Buunen et al. 1974; Cabot et al. 1976; Goldstein, 1967).

Goldstein (1967) further explored differences between AM and QFM signals and reported that QFM signals were perceived as being "smoother and steadier" than AM signals for low modulation frequencies. Quality differences between the two signals diminished with increase in modulation frequency. Since phase was the only differing parameter between AM and QFM, the implication is that "phase effects disappear for stimulus bandwidths that exceed a value roughly proportional to the critical band at the carrier frequency" (Goldstein, 1967, p.458). Well-resolved components would thus be less susceptible to phase effects than poorly-resolved components.

The relation between roughness and spectral filtering was also observed by Terhardt (1974). For sinusoidally amplitude-modulated tones (SAM), with carrier frequencies ≤ 1000 Hz, maximum roughness was observed for modulation frequencies < CB. The degree of modulation 'm', (equivalent to the relative fluctuation amplitude) of AM signals was found to be related to roughness 'r' via the relation r= const. m^2.

Terhardt (1974) attempted to **"scale"** roughness using a magnitude estimation procedure in which subjects were required to estimate the perceived roughness of a test signal to be "more or less than half as rough", or "twice as rough" as a comparison signal with fixed 'm'. The results indicated that listeners were able to make such judgments

reliably. The ratio of the test-signal modulation depth to that of the reference (m_t/m_{ref}) was found to be independent of carrier frequency and reference modulation depth m_{ref}. The ratio m_t/m_{ref} in all conditions appeared to be =0.707 for estimates of "half as rough", verifying the relation proposed earlier by Terhardt; namely that roughness is proportional to the square of the degree of modulation.

Additional experiments led Terhardt (1974) to conclude that "the entire roughness" of a complex sound "is composed of partial roughnesses which are contributed by adjacent critical bands". In order to avoid roughness, therefore, he advises that "high harmonics have to be attenuated or even suppressed. The amplitudes of the present harmonics have to be chosen in such a way that as few harmonics as possible fall into one and the same critical band" (p.212).

The sensation of roughness arising from the interference of harmonics has come to be regarded as "an integral part of the timbre of wide-band stimuli" (Plomp, 1976, p.97). Timbre depends, thus, not only on the spectral envelope of a sound, but also on the interference of spectral components. Such interference, in turn, is dependent on spectral spacing of components. Sounds of high spectral density are more susceptible to phase effects. The discrimination of spectral density is "primarily based upon the perception of temporal fluctuations in the intensity of the sound, and secondarily upon resolved structure in the spectrum, perceived as tone color" (Hartmann et al., 1986).

Timbre may thus be considered as being dependent on at least

two aspects of sound spectrum: the *absolute* position or "**spectral locus**" of components, and their *relative* position, or "**spectral spacing**" (Singh, 1987).

2.7.5 Dimensional analyses and perceptual scaling of timbre

Difficulties in quantification of a multidimensional perceptual attribute like timbre were alluded to in the introduction of section 2.7. While a simple measure has not yet presented itself for universal application in description of timbre differences, a few undaunted attempts to "scale" timbre have met with some success. These investigations have typically used a comparative approach, whereby sounds are rated individually or relative to each other along some scale. Such ratings have involved strength of allocation of verbal attributes (von Bismarck, 1974), judgments of perceived similarity (Grey, 1975, 1977; Plomp, 1970; Wedin and Goude, 1972) and the finding of perceptual "analogies" (Ehresman and Wessel, 1978).

2.7.5.1 Adjectival descriptors of timbre

i) Semantic scales

Nominal labels characterizing properties of a sound source or qualities of the sound produced have always pervaded as descriptors of

timbre differences. Verbal adjectives such as "bright", "full", "dull", "rough", "open" have been used for aeons in music and indeed, in everyday perceptual experience with complex sounds.

Von Bismarck (1974 a) attempted to quantify such usage of verbal attributes of timbre by presenting 35 different sounds to listeners, and asking them to rate them along 30 different semantic scales. A "semantic scale" or "semantic differential" was basically a 7-point scale bounded by a pair of polar adjectives such as "dead - lively", "thin - thick", "strong - weak", "sharp - dull" etc. Of the 35 sounds, 5 were noise bands, while 30 were complex tones with F0=200 Hz and different spectral slopes and densities.

A principal- components analysis of the vector space derived from correlation coefficients of the 30 scales revealed four dominant factors that accounted for 91% of the variance. Of these, the two most important factors appeared to be related to the scales contrasting "sharpness" with "dullness", and "compactness" with "scatter". The latter scale was clearly used to differentiate tonal stimuli from noise. The former scale appeared to be related primarily to the center of gravity of the sound spectrum rather than to spectral- envelope shape. It alone accounted for 41% of the variance in the rating data.

ii) "Sharpness"

Given the importance of the **"sharpness"** scale in the study of verbal attributes, von Bismarck (1974 b) went on to explore the relation

between perceived sharpness and physical parameters of sounds in another study. The physical variables manipulated were: the upper and lower limiting frequencies of the spectral locus (between 200 Hz and 16 kHz), spectral slope (-6, 0, +6 dB/oct) and F0 (100, 200, 400 Hz) of the sounds. A magnitude estimation procedure of the type employed in scaling pitch and loudness was used (Stevens et al., 1937; Stevens, 1955).

The results basically showed that sharpness too could be doubled and halved. Further, it appeared to **increase with the upper frequency limit** of the spectrum and with rising spectral slope. Sharpness judgments for sounds of differing F0 and sound-pressure level were more variable, but it seemed that sharpness was **"an attribute distinguishable from pitch and loudness"** (p. 159).

The "formation" of sharpness appeared to be "characterized by the combined effects of: (1) the position of energized spectral regions, and (2) the magnitudes of energy in those regions" (p. 169). Sharpness thus appears to be closely related to the attribute **"brightness"** studied by Lichte (1941) which was also found to depend on the location of the center of the energy distribution along the frequency continuum.

2.7.5.2 Multidimensional scaling of timbre differences

Given the multidimensionality of timbre and lack of a common standard by which to compare timbre changes, Plomp and Steeneken (1969) and Plomp (1970) resorted to an experimental procedure

whereby sounds could be compared in terms of their *relative* similarity to each other. The similarity data thus obtained were analyzed via a data-processing technique specifically designed to probe and seek out patterns embedded in the data. This technique (referred to as "multidimensional scaling" or MDS), allows the data structure to be represented spatially in the form of a geometric model. Distances between points in the spatial representation are indicative of the similarities (or differences) between the stimuli compared. Comparison of dimensions of the perceptual space thus derived and the physical space derived from stimulus differences then allows physical determinants of the perceptual judgments to be inferred. [Details about the development and application of such MDS techniques can be found in a volume of papers edited by Shepard et al. (1972)].

i) Use of MDS techniques to study the effects of phase on timbre

Given the surprising scarcity of previous studies that **directly** investigated the role of phase in the perception of timbre, Plomp and Steeneken (1969) applied MDS techniques to investigate the suspected dependence of timbre on phase quantitatively.

A method of "triadic" comparisons was used that entailed pairwise rating of similarity of three sounds available per trial. These were drawn from a stimulus set comprising 8 stimuli with equal f0 and amplitude relations of harmonics. The only difference between the

sounds was in their phase relations and thus in their waveforms. All the $({}^{n}C_{3})$ triads obtained from 3-way combinations of n stimuli were judged in this way and a matrix of "dissimilarity" indices derived. An MDS algorithm was applied to the matrix to compute a Euclidean timbre space in which linear distances between two points corresponded to their timbre dissimilarity.

Despite differences in waveform, the stimuli were perceived as being very similar. The maximal difference between two sounds appeared to occur when the phase of alternate components of the respective harmonic series were 90 degrees out of phase as in the case of a series with alternate sine and cosine components being compared with a series of only sine or only cosine terms. This maximal inter-point distance in the timbre space was preserved even when other stimuli with randomly selected phase spectra were included in the stimulus set.

A dependence on f0 was noted however, with stimuli of low f0 being more susceptible to phase effects than those of higher f0. As noted before (in section 2.7.4), this observation is not surprising, given the more densely populated spectrum of tones with low f0. Interaction of components lying within critical bands would thus be more likely, particularly for the higher components of low-f0 tones.

Plomp and Steeneken thus concluded that the ear was **not insensitive to phase differences**, but that the timbre changes accompanying phase changes appeared to be small in comparison with the influence of amplitude spectrum on timbre.

ii) Use of MDS techniques to study timbre and the amplitude spectrum

Plomp (1970) also applied MDS techniques and the method of "triadic" comparisons to study the dependence of timbre on amplitude spectrum. The triads were drawn from a stimulus set comprising the "steady-state" portions of 9 complex tones sampled from real instruments. The timbre space derived from the dissmilarity indices was compared to the spectral space derived from spectral differences between the tones. This latter physical space was obtained by comparing differences in sound-pressure level across 15 1/3 octave-band filters for pairs of tones taken from the stimulus set.

Such an analysis was undertaken because of its inclusion of features of the hearing mechanism, in that the outputs of the filters could be considered to be analogous to the distribution of loudness across critical bands. Correlation coefficients of the two 3-dimensional configurations obtained for the two spaces were very high (~0.91 to 0.99). After all this manipulation, Plomp arrived at the (foregone) conclusion that relative dissimilarities in timbre can indeed be predicted quite well from differences in sound spectrum !

2.7.5.3 Investigation of timbre by analysis/synthesis

The analysis-synthesis paradigm mentioned briefly in section

2.7.1, has proven to be a powerful tool with which to "explore" the multidimensionality of timbre (Grey, 1975; Risset and Wessel, 1982). Its appeal is further enhanced by the incorporation of both physical and perceptual determinants of timbre, and potential utility in enabling parsimonious data reduction in synthesis schemes. The basic idea behind the analysis-synthesis approach is recapitulated in **figure 2.6** (adapted from Risset and Wessel, 1982).

The analysis stage basically entails the estimation of physical parameters (or "control functions") from the original sound. These can then be used as input parameters for the synthesis stage. A perceptual comparison of the original and synthetic versions of the sound enables the unveiling of parameters that are the most important or salient determinants of timbre.

The analysis-synthesis approach, coupled with MDS techniques, was adopted by Grey and associates at Stanford University in work on "multidimensional perceptual scaling of instrumental timbres" (Grey, 1975, 1977, 1978; Grey and Gordon, 1978; Grey and Moorer, 1977). In addition to consolidating important insights about the underlying determinants of timbre, the *techniques* developed in the course of their research have provided useful tools for future studies of timbre.

Using a heterodyne-filtering analysis technique developed by colleague Moorer (1974), Grey analyzed 16 tones generated by natural instruments (brass, strings and woodwinds).

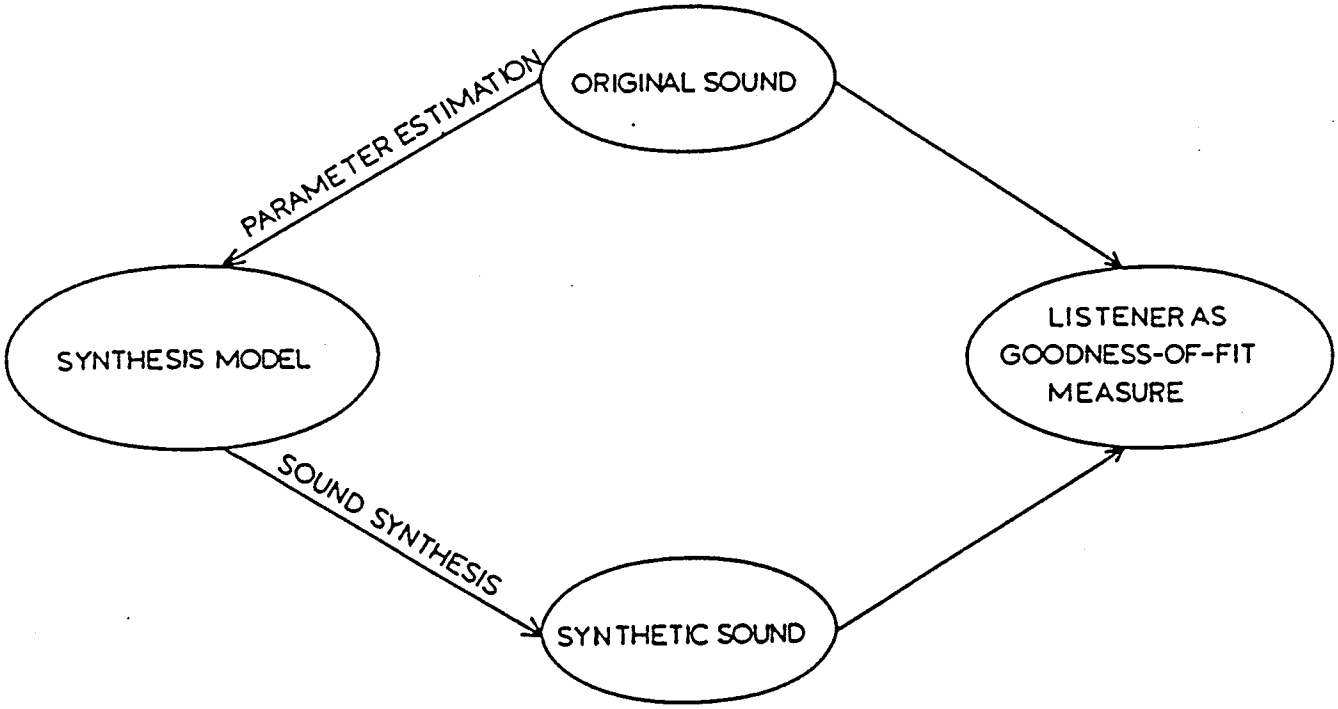


Figure 2.6 Conceptual framework of the analysis-synthesis approach. The sound to be modeled is first analyzed. The parameters estimated from the analysis are used to generate a synthetic version of the original sound. Auditory comparisons between the original and synthetic versions of the sound serve as a measure of the "goodness - of- fit" of the synthesis scheme. If discrmination is poor, essential features of the sound are considered to have been retained through the analysis, data-reduction, and synthesis stages. [Adapted from Risset and Wessel, 1982].

This heterodyne technique (similar in principle to the "phase vocoder" described by Flanagan and Golden, 1966) yielded time-variant amplitude and frequency/phase functions for each component of each complex tone. The analysis data were used in different ways in the synthesis process: In one case, **all** the information available was used to generate a more or less exact replica of the original tones (a little degraded by processing "noise" perhaps). In other cases, the complex time-variant functions obtained for individual components of the tones were **simplified** using different approximation schemes. The simpler, "data-reduced" functions were then used in the synthesis process.

The results of a discrimination experiment using the original and synthetic tones were reported by Grey and Moorer (1977). Of three different simplification schemes used for synthesis, they found a "line-segment" approximation of the time-varying frequency and amplitude control functions to be the most successful (i.e. the **least discriminable** from the original). **Figure 2.7** illustrates the time-variant spectra of original and synthetic versions of a bass clarinet tone, where the latter was synthesized using the "line-segment" approximation.

The success of this modification has sometimes been taken to imply that many of the fluctuations in the complex microstructure seen in dynamic spectral analyses are not that essential for the perception of timbre. This observation may seem a little paradoxical, given the

importance of transients established previously, and the important role of modulations in perceptual fusion. While some of the fine details of the microstructure may indeed be dispensable, it should be noted however, that the line-segment approximation used by Grey and Moorer *did* in fact preserve some information about the time-variant amplitude and frequency behavior of spectral components. The extent of this dynamic information was apparently sufficient in conveying timbre, and presumably, in maintaining the fusion of components into a *gestalt* tone.

A further need for caution against over-generalizations stating that fluctuations in frequency and amplitude are unimportant is indicated by the fact that two other approximations used by Grey and Moorer led to loss of timbre similarity (i.e. improved discrimination). These were the "constant-frequencies" and "cut-attack" approximations. The former preserved information about variation of amplitude, but not frequency, and the latter ignored some information coded in the initial "attack" segment of the sound. The loss of these details led to the synthesized tones being perceived as being dissimilar to the original versions.

Given the success of synthetic tones in conveying the original timbral identities of instrument tones, Grey carried out further experimentation with data-reduced tones as stimuli. The fact that they were generated via computer allowed greater flexibility in manipulation of specific parameters, the relative perceptual importance of which could then be studied.

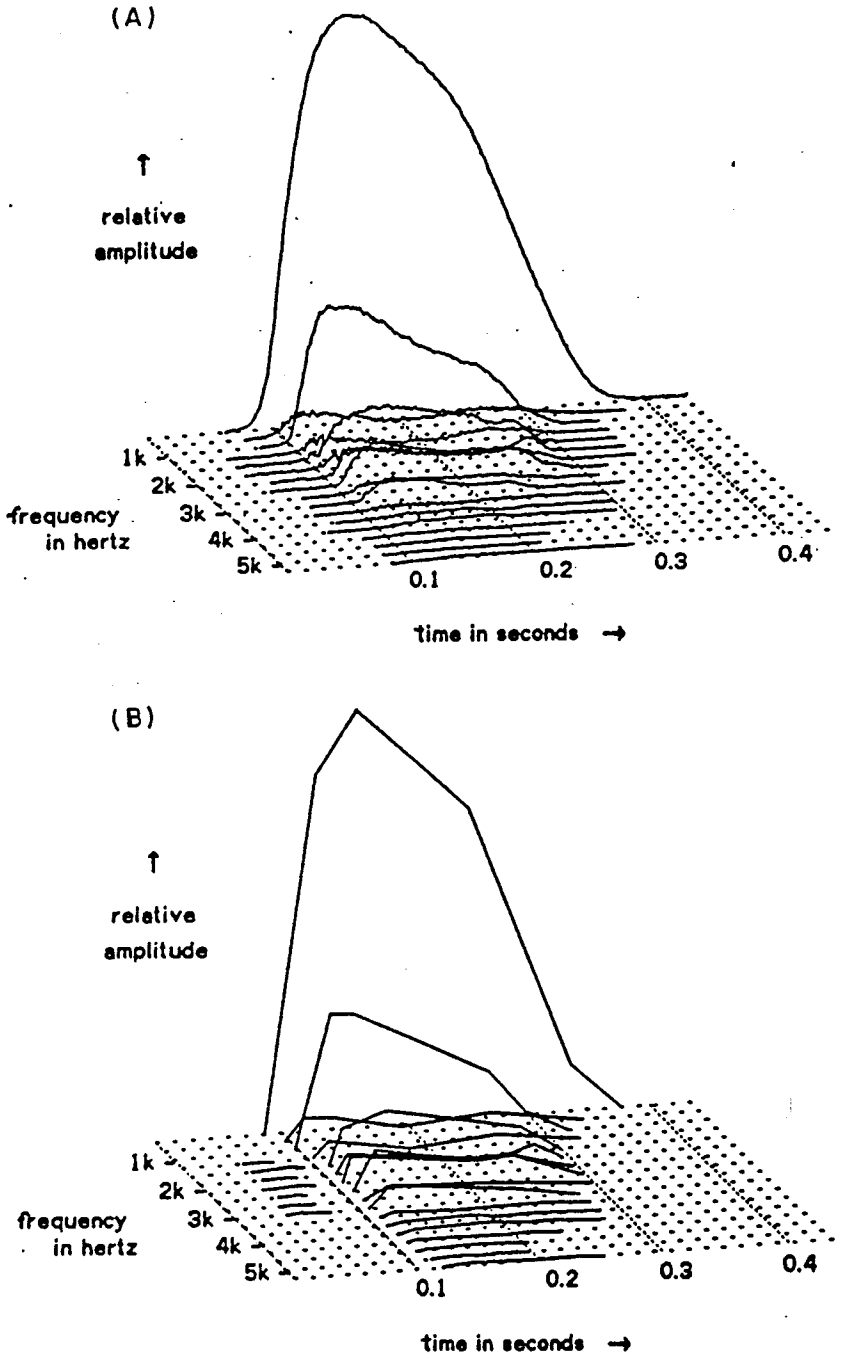


Figure 2.7 Analysis data used in synthesis of a brass clarinet tone shown in a 3-dimensional perspective (axes = frequency, time and amplitude). (A) shows time-variant amplitude functions derived from a heterodyne analysis. (B) shows "line-segment" approximations of the functions in (A). Considerable data-reduction was achieved by the approximation. [From Grey and Moorer, 1977]

In one such experiment,16 synthetic tones were presented to listeners for pairwise comparison of perceptual similarities. Dimensional analysis of the "timbre space" obtained by MDS of the dissimilarity data confirmed the **importance of the distribution of spectral energy** as a major factor distinguishing timbres. Another major dimension defining the space related to temporal factors such as the relative synchronicity in onset of harmonics and the transient spectral distribution during the "attack" portion of sound initiation. While the former dimension is related to "brightness" and "sharpness" (von Bismarck,1974 b; Lichte, 1941), the latter temporal dimension has been identified by Grey (and by Wessel, 1973) as one enabling a set of timbres to be perceived as belonging to the same instrument "family".

2.7.5.4 Conceptual and perceptual navigation in a timbre space

The successful representation of perceptual similarities of timbre in a geometrical space has opened up a whole new realm of possibilities for the study and creative use of timbre.

The results of Grey (1975), Moorer (1977), and Grey and Gordon (1978), provided evidence that a timbre space could be used to create changes in parameters of synthetic tones to yield perceptual changes that were consistent with the geometry of the space. Such changes were provided, for example, by the exchange of spectral envelopes between

tones, or by using interpolation algorithms to create tones lying between chosen end-points in the space. In the former case, exchange of spectral envelopes between tones led to a corresponding shift in the location of the tones in the space. In the latter case, the tone underwent a timbral transition, starting with the timbre of the original tone and ending with the timbre of the final tone via a series of in-between timbres. If the end-point timbres were similar to begin with, the transitions appeared smooth. If they were very different, the intermediate sounds appeared "strange", "sometimes sounding like more than one instrument is playing" (Moorer, 1977, p. 1117).

The fact that the coordinates defining the timbre space could be manipulated to yield novel or transitional timbres, led Wessel and colleagues to further explore the navigational viability of timbre spaces in terms the perception of **relations** between timbres and timbral "trajectories" in the space (Ehresman and Wessel, 1978; Wessel, 1979/1985; Wessel et al. 1987). It appears that analogical relations exist for timbre patterns that could enable the "transposition" of timbre intervals in a manner similar to the transposition of pitch intervals. It also seems potentially viable that timbre variations may be quantifiable in terms of graded differences in the space.

These exciting observations of Wessel et al. came about in the course of an extensive effort to develop controls for manipulating sound structures in computer music. One focus of their effort has been the determination of "constraints" that would need to be imposed on a timbre

space to maintain perceptual constancy of timbre across such variables as pitch registers, dynamic markings, and phrasing and articulation in melodic patterns. While providing a dazzling glimpse of ways in which timbre can function, both as a perceptual attribute and as an organizer of melodic material, the information gleaned from these preliminary studies is largely heuristic in nature.

A more rigorous experimental approach coupled with these new techniques of sound production and control holds great promise in providing a cohesive understanding of the psychoacoustics of timbre and auditory processing of complex sounds in general.

2.8 Pitch perception and mechanisms of hearing

In contrast to the paucity of research directly addressing the issue of timbre perception, the perception of pitch has received consistently greater attention. The relative ease of comprehension of the nature of the pitch percept by most listeners, and the **fundamental** role of pitch in music have rendered it an important and approachable percept to be examined in psychoacoustic experimentation.

In addition to being investigated for its own sake as a major perceptual attribute of sounds, pitch has also served a useful, adjunct role as a **tool** in auditory research. Pitch is related to the frequency of simple tones and to the F0 of harmonic complex tones. Since frequency can be fairly well-controlled in sound production and transmission, it is often

used as an experimental variable in studies of the hearing mechanism. Information about underlying auditory processes can then be gleaned by noting changes in pitch associated with changes in frequency.

Several issues have pervaded in research *on* and *with* the pitch percept. Some of them pertain to pitch perception at a basic level in terms of stimulus features and the encoding of these features by the auditory system. Others pertain to a higher, more cognitive level of pitch perception, whereby pitch material is organized into larger perceptual structures such as intervals, chords and melodies.

Some of the more basic questions about pitch processing that have arisen in the course of research on psychoacoustics are listed below:

1. Why does a group of simple tones fuse together to yield a complex tone with a single pitch ?
2. Why is this unitary pitch usually equivalent to the frequency of the fundamental component even when the latter may be physically absent in the spectrum ?
3. Why does a single pitch often predominate even for inharmonic complexes lacking a common fundamental ?
4. Is it frequency per se, or the ***periodicity*** of a complex that is responsible for the perception of a unitary pitch ?
5. What are the mechanisms responsible for the encoding of pitch (whether in terms of frequency or periodicity) in the auditory system ?
6. At what level of the system do these mechanisms operate ?

Some of these questions have been answered in the course of the last century, while controversy and confusion continue to surround others. General factors that enable perceptual grouping of simple components to form a fused complex sound have already been discussed in section 2.4. The issue of assignment of a single pitch to a fused tone and other insights gained about pitch perception are reviewed in the following sections.

2.8.1 A "place" theory of pitch perception

In addition to establishing the relation between pitch and physical frequency, von Helmholtz (1877/1954) claimed that frequency was encoded in the ear in terms of regions of maximal stimulation on the basilar membrane. A **"place" theory** for pitch perception was thus proposed: Since pitch was related to fundamental frequency, and frequency was related to place, it followed that pitch was related to place of stimulation.

The idea of a such a "spatial" frequency analysis, or **"tonotopic" organization of frequency** in hearing was corroborated to some extent by the experiments of von Bekesy (1928/1960). Von Bekesy did not find the type of tension properties necessary for the basilar membrane to respond to different frequencies as selectively as suggested by von Helmholtz. He observed instead, that a stiffness "gradient" existed along

the length of the membrane, that led to the creation of "travelling waves".

The peaks of the envelopes of these waves were located at different places along the membrane, with high frequency stimuli corresponding to "maxima" toward the basal end of the cochlea, and low frequency stimuli corresponding to maxima toward the apical end. The observed interaction between place of maximal stimulation and frequency of exciting stimulus, lent strong support to the idea of a "place" principle for the perception of pitch.

2.8.2 "Fundamental" challenges to Ohm and von Helmholtz' conception of pitch

The simple relation established by von Helmholtz between pitch and place and F0 was challenged by Seebeck (1843). Using an acousical "siren" with which the periodicity of recurring puffs of air could be controlled, he demonstrated that it was the **period** of the signal waveform, rather than the frequency of the lowest Fourier component that corresponded to the perceived pitch of the complex sound generated by the siren. His statement of his findings reads:

" ... a tone with pitch m is not necessarily of the form $a\cos 2\pi(mt+\phi)$ but it may contain also terms of the form $a_s\cos 2\pi(smt+\phi_s)$, where s can be very large and a_s very small. It should not even be

excluded that those latter terms, when taken together, can evoke a tone with pitch m when a term of the form $a\cos 2\pi(mt+\phi)$ is not present..." (translated by de Boer, 1976, p. 502).

This remarkable hypothesis contradicted "Ohm's acoustical law", claiming that the pitch of a complex tone could be conveyed by high harmonics and could even equal that of a tone **absent** in the spectrum. However, due to limitations in alternative procedures for verification of the absence of the F0 at the time, and the antagonistic response of the venerable Ohm and von Helmholtz, Seebeck's "period" challenge was virtually ignored and discounted.

Other researchers subsequently made observations similar to those of Seebeck, but reconciled their results with Ohm's law by suggesting that in cases where the fundamental component was absent in the physical stimulus, it was later reintroduced as a "difference tone" by nonlinear distortion in the hearing organ (Everett, 1896).

With the development of telephony in the early part of the twentieth century, **"the case of the missing fundamental"** was reopened. Telephone circuitry was bandlimited in frequency and restricted the transmission of frequencies corresponding to the fundamental frequency of the human voice (<300 Hz). Despite this physical attenuation of low frequencies, the perceived pitch of the transmitted voice appeared to be unchanged ! This curious phenomenon was explained by Fletcher (1924) in terms of the distortion hypothesis invoked by his predecessors. He claimed that despite its absence in the

physical stimulus, the fundamental component was reintroduced in the ear via a non-linear process that resulted in the generation of distortion products, such as difference tones. If the transmitted components were harmonic multiples of the missing fundamental, then the unchanged pitch could be explained easily as being equivalent to that of the difference tone, which was equivalent to the fundamental frequency.

This interesting technical development thus did not lead to any immediate, new, insightful developments in pitch theory. It did however reignite the spark of interest in the claims made by Seebeck and kept the idea of periodicity being the "fundamental" contributor to the pitch of complex tones alive and debatable !

2.8.3 A lack of resolution

Schouten (1938, 1940) replicated Seebeck's experiments, demonstrating the unchanged pitch of a complex tone, despite elimination of F0 in the spectrum. The observation that the waveform generated by the residual higher components retained the original periodicity led Schouten to claim that:

"the lower harmonics can be perceived individually and have almost the same pitch as when sounded separately. The higher harmonics however, cannot be perceived separately but are perceived collectively as one component (the residue) with a pitch determined by the periodicity of the collective waveform, which is equal to that of the

fundamental tone" (cited in Plomp, 1976, p.112).

This low pitch, termed the **"residue",** was defined as "that subjective component of sound sensation that is the result of the combined impression of higher harmonics, namely those harmonics that cannot be resolved by the auditory organ"

Schouten's residue hypothesis set the proverbial "ball" rolling in the field of pitch research. The residue theory provided a reasonable, alternative explanantion for the case of the missing fundamental without invoking the idea of distortion processes. It also triggered a fresh debate on the subject of frequency coding in the auditory system.

2.8.4 "Temporal" theory of pitch perception

The "place" theory of frequency coding proved inadequate in accounting for the phenomenon of the missing fundamental. With distortion being dismissed as an explanation, doubts raised earlier about place coding for "interruption-tones" and the pitch of noise interrupted at different rates also resurfaced (Miller and Taylor, 1948).

Wever (1949) reiterated the idea of a temporal mechanism (suggested earlier by Rutherford (1886) and Wever and Bray (1930)) for encoding stimulus frequency to explain such phenomena that could not be accounted for by a place theory. He proposed the **"volley" theory** of hearing, suggesting that auditory neurons responded in **groups** to stimulation by a periodic stimulus. Instead of a single neuron

discharging periodically at a rate corresponding to the stimulating frequency, it was hypothesized that groups of neurons could fire at phase-locked intervals derived from the stimulus waveform. The net firing pattern would thus be able to convey the frequency of the stimulus and the timing information could be used by the central auditory system to derive the associated pitch.

The volley theory came to be viewed as a serious contender in explaining how the auditory system perceives pitch. The fact that **both "place" coding and "temporal" coding** theories could account for pitch perception to differing extents, led to a dilemma that manifested itself experimentally in a number of studies that followed.

2.8.5 Conflicts of interest

An experiment by Davis et al. (1951) highlighted the type of conflict that can potentially occur if place and period cues give different information. Their listeners were required to match the pitch of "tone pips" produced by brief rectangular pulses passed through a filter about an octave-wide, centered at 2000 Hz. The pulsing frequency varied from 90-150 Hz. Listeners had a difficult time making pitch-matches, sometimes choosing to match the pulsing frequency and sometimes choosing to match the 2000 Hz frequency. Davis et al. colorfully referred to the former pitch as the **"buzz"**, and the latter as the **"body"**. A lot of listeners were confused by the multiplicity of potential matches and

could not do the task until aided by some guiding factor such as a dynamic change in the frequency of either the buzz, or the body. Davis et al. thus argued that pitch is a "**double attribute** compounded of "buzz" (correlated with frequency of volleys of nerve impulses) and "body" (correlated with position of maximal stimulation on the basilar membrane)".

This idea of pitch being a "double" attribute was further formalized by Licklider (1954). He conclusively demonstrated the difference between **"place" pitch**, and **"periodicity" pitch** by experimentally showing that the "low" pitch of a harmonic complex tone lacking a fundamental component was derived from the temporal "period" of the stimulus, rather than a place of maximal stimulation corresponding to some "pseudo" fundamental reintroduced by distortion. He used masking noise located either in the region of the fundamental, or in the region of the higher components and showed that the former type of noise did not succeed in masking the low pitch, while the latter noise did. Thus, although the pitch was matched to the low, absent fundamental frequency, it appeared to be dependent on some feature of the present, high harmonics, presumably, the period of their combined waveform.

The existence of both place and periodicity as viable principles to encode frequency, were reconciled by Licklider (1955) in a **"duplex theory"** of pitch that incorporated features of both these principles. (A concise review of this theory can be found in de Boer, 1976).

2.8.6 Pitch as a double attribute: A musical viewpoint

Duality in the nature of the pitch percept has also been noted in music. "The intuitions of musicians ... suggest that there are at least **two dimensions** to musical pitch" (Dowling and Harwood, 1986, p. 107). These have been described as **"tone height"** (or "pitch height"), and **"tone chroma"** (or "pitch class") by many music theorists (Bachem, 1950; Revesz, 1954; Forte, 1973).

These two dimensions were deemed necessary to describe musically significant pitch relations such as the perceived similarity of tones separated by octave intervals (F0 ratio=2:1) and the transformational invariance of melodies under transposition (displacement of absolute F0's, while preserving F0 ratios between successive notes).

In the basic model developed to incorporate these two dimensions, pitch is represented by a **helix** (illustrated in **figure 2.8**). Helical motion comprises both rotation and translation along the axis of rotation (Shepard, 1982). The translational or "rectilinear" component corresponds to "tone height" and allows sounds to be distinguished in terms of octave-type relations, while preserving the note name or "chroma", which corresponds to the circular dimension of the helix.

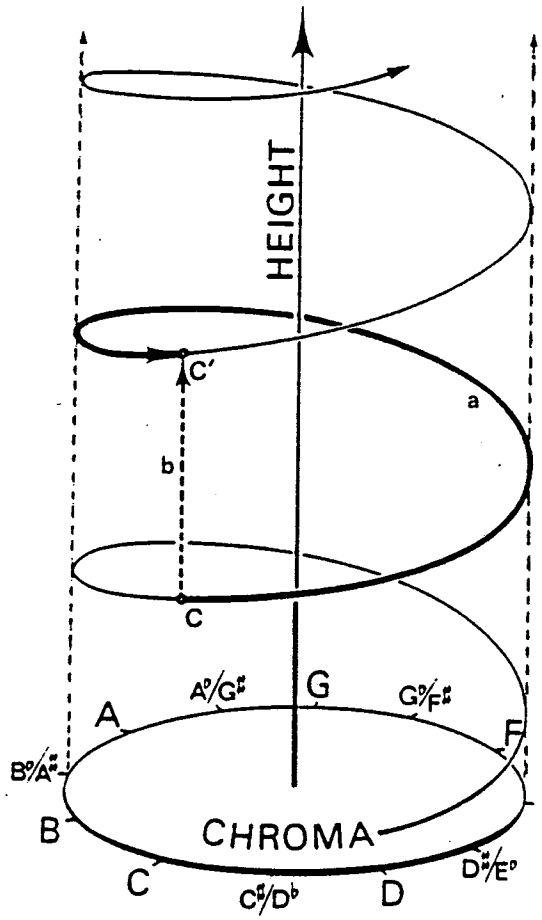


Figure 2.8 Helical representation of musical pitch. The vertical dimension corresponds to perceived "tone height", while the circular dimension corresponds to "tone chroma". Perceived similarity between two notes C and C' separated by an octave, is observable in the model by the reduced distance of separation when measured along the linear axis (path 'b'), as opposed to the greater distance along the circular dimension (path 'a') [from Shepard, 1982].

Notes with F0 differences corresponding to successive steps of a "chromatic" scale spanning an octave would occupy successive positions along the circular dimension. One complete rotation would thus amount to a progression of ascending or descending pitches converging at the same "note" separated by an octave.

Two notes C and C' with an F0 ratio equal to an octave (1:2) represented on such a helix, would have the same "chroma", but the latter would have greater "tone height" (see figure 2.8). Perceived similarity between these tones would be observable in the model by the reduced distance of separation when measured along the linear axis (path 'b' in figure 2.8), as opposed to the greater distance along the circular dimension (path 'a').

The perceptual similarity of notes separated by pitch intervals corresponding to an octave often renders them functionally equivalent in music. While there is some rethinking about harmony in the current music-theoretic community, the inversion of notes in chords has traditionally been considered as not changing the "harmonic" function served by the chord (i.e. a **simultaneous** combination of notes C, E and G is conventionally taken to be equivalent to the combinations E-G-C', G-C'-E' etc.). In melodies, however, the *succession* of pitches is the defining factor and "their order becomes most essential" (Toch, 1948/1977, p.63). A **sequence** of notes C-E-G would be perceived as being a totally different melody than that articulated by the sequence G-C'-E.

Given the dichotomy of tone height and tone chroma, conflicts in

pitch perception (similar to the type reported by Davis et al.,1951) can arise when divergent information is provided in an acoustic stimulus.

Risset (1971, 1986) has reported many "pitch paradoxes" that arise when listeners are presented with a sequence of tones that ascend along one dimension of pitch (such as height), while simultaneously descending along the other dimension (chroma). Listeners can be confused by conflicting pitch cues of this type, an observation that was particularly highlighted in an experiment on "octave generalization and tune recognition" by Deutsch (1972). She found that the scrambling of notes across octaves while preserving their chroma led to even familiar melodies becoming unrecognizable. The changes in pitch height implied a different melodic contour than did the changes in chroma. Preservation of "contour" is a critical factor in melody perception (Dowling and Harwood, 1986).

The relation of tone height and tone chroma, to acoustical stimulus dimensions will be discussed further in a later section as well as in the next chapter. For the moment, it should simply be noted that pitch is not, in fact, a simple, unidimensional auditory attribute **"in terms of which sounds may be ordered on a scale extending from low to high"** as the ASA (1960) definition would imply. Rather, there are *at least two* dimensions of pitch: one corresponding to absolute frequency positioning of the stimulus and the other to more relative aspects of the spectrum, such as the spectral spacing of components (equivalent to F0 for harmonic tones).

2.8.7 The low pitch of inharmonic complexes

The confusion between place pitch and periodicity pitch observed by Davis et al. (1951) was explored further by Thurlow and Small (1955) using repeated pulse-train stimuli. Such stimuli can be regarded as "residue"-type stimuli comprising high harmonics of an absent fundamental when the repetition frequency is a multiple of the pulsing frequency. Stimuli in which the interrupted frequency is not a multiple of the repetition frequency no longer maintain a harmonic relation between the partial components. A low pitch was reported by Small (1955) for these type of **inharmonic** stimuli as well.

De Boer (1956/1976) also reported results of a number of experiments using inharmonic signals. His curiosity about such signals was apparently aroused by the report of Schouten (1940) that shifting the frequencies of all harmonics of a complex signal by an equal amount of Δf (in Hz) led to a shift in the low pitch of the complex. This was an intriguing situation since *neither the place, nor the period* of stimulation appeared to correspond exactly to the frequency matched for the new pitch !

De Boer studied this type of inharmonic situation further, using amplitude modulation to generate complex signals with 5 or 7 components that were equally spaced in frequency. The spacing equalled the modulation frequency ("g") and the frequency of the middle component equalled the carrier frequency ("f"). The signals thus comprised a series of components with frequencies:

f-3g, f-2g, f-g, f, f+g, f+2g, f+3g

If f=ng, this type of stimulus is harmonic and the perceived pitch is typically matched to the missing F0 which equals the spectral spacing (equal to the modulating frequency g) of the harmonic components. If the carrier frequency is changed from f to f+Δf, so that f≠ng, the stimulus becomes inharmonic. In this case, the components are no longer multiples of a common F0 (=g), but the period of the resultant waveform remains unchanged.

If a "period" principle was in strict operation, the pitch of the latter, inharmonic complex should have remained the same, given the unchanged waveform envelope. But de Boer observed (as had Schouten), that **the low pitch changed** in the direction of change of the carrier frequency. This proportional relation between perceived pitch and the center frequency can be represented as $\Delta p=\Delta f/n$, and has been referred to as the **"first effect"** of pitch shift (de Boer, 1976). The change in pitch, however, was not strictly proportional to the carrier frequency (i.e. the frequency of the central component). Large deviations were found at low f/g ratios. With an increase in the spacing g, keeping the center frequency f constant, the pitch of the complex appeared to go down. This type of deviation from proportionality was considered to be a "perturbation" by de Boer and called the **"second effect"** of pitch shift.

To explain the observed pitch shifts, de Boer proposed the idea that the auditory system does not respond simply to the gross temporal envelope as suggested by the periodicity principle, but is also capable of

picking out approximate periodicities from the **"fine structure"** of the waveform. The matched pitch would then correspond to the inverse "pseudo period" estimated in this way.

De Boer also found that a residue pitch was perceived for signals comprising components that ought to be well-resolved by the auditory system. Further, the pitch of such a complex underwent a similar pitch shift when the components were shifted in frequency by a linear amount of Δf. He thus made a similar modification of the "place" principle and proposed a "pseudo fundamental" theory. According to this theory (described in de Boer, 1976, p.519), "the pseudo-fundamental is equal to an integral submultiple of the carrier frequency" to a first approximation.

When a signal is inharmonic, "the auditory system tries to find a harmonic series of frequencies that corresponds in the best possible way with the frequencies of the components presented". He also suggested that in finding the "best fitting" pseudo-fundamental, the auditory system may assign different weights to different components. Since the lower components of a complex undergo a larger "relative" shift when the frequency of components is altered by the same linear amount, they may be assigned a higher weighting in the pitch estimation process.

2.8.8 Differential contribution of components to the overall pitch of a complex

Unequal weighting for different components was investigated in

two contemporaneous experiments by Plomp (1967) and Ritsma (1967). They generated stimuli that comprised harmonics of two different F0's. In a paired comparison of such stimuli, a conflicting situation was presented to listeners: One set of harmonics was incremented in one direction, while the other was decremented. The reported direction of perceived pitch change was taken to imply that the corresponding harmonic or set of harmonics was more "dominant" in conveying the pitch of the complex. In both experiments, the lower, better-resolved components were found to be more dominant in indicating the change in pitch. Ritsma further narrowed down this dominant spectral region to be that encompassing the third, fourth and fifth harmonics for the range of F0's tested (100-400 Hz).

The original definition of the residue claimed it to be the low pitch evoked by the joint perception of a group of unresolved **high** harmonics. The results of Plomp and Ritsma showed instead that to a fairly large extent (≈fo=1000 Hz), it is the **low** components (though not necessarily the fundamental component) that dominate the perception of pitch. De Boer (1976) thus suggested redefining the residue as being simply the "joint perception of a number of components" dominated in several cases by the lower harmonics in the spectrum.

It also appears that the "dominance region", may in fact be a straightforward consequence of basic auditory sensitivity to sounds of different frequency. The 500-2000-Hz dominant range obtained in the experiments of Plomp (1967) and Ritsma (1967) support this notion. This frequency range does fall within the region of high sensitivity on the

audibility curve (Robinson and Dadson, 1956). Plomp's later writings (Plomp, 1976, p.142; Rasch and Plomp, 1982, p. 9) also indicate that the dominance region is probably better defined in terms of absolute frequency, rather than in terms of relative frequency relations (as expressed by harmonic numbers). Thus, the low pitch of complex tones with low fundamental frequencies (under 500 Hz) depends on the higher partials (falling in the dominant region). The harmonic numbers of the dominant region decrease gradually for increasing F0, reaching the fundamental itself for F0~2000 Hz.

2.8.9 Combination tones: The invisible spectrum

Smoorenburg (1970) wished to test the range of applicability of the pitch shift phenomena reported by de Boer (1956/1976) and Schouten et al. (1962) for stimuli that comprised only **two** frequency components. He obtained results similar to the earlier pitch shift data when component frequencies were linearly shifted from harmonic values. However, the steepness of the deviation function implied that components lower in frequency than those present in the stimulus were being used to determine pitch.

Based on these observations, Smoorenburg suggested that the pitch was derived from combination tones introduced by the non-linearity of the ear. The frequencies of such combination tones can be described by the relation $f1 - k(f2-f1)$, where $f1<f2$, $f1/f2$ is not too large and k is a

small integer (Goldstein, 1967). These "invisible" components correspond to lower stimulus frequencies in both harmonic and inharmonic situations. A simultaneous increase of component frequencies f1=ng and f2=(n+1)g by Δf would give combination tones of the type (n-k)g+Δf that were similarly shifted in frequency by Δf. The large pitch shift obtained could then be explained as ensuing from this shift in the frequency of combination tones in a region lower than the the components of the original stimulus.

Combination tones were also considered to be the explanation for the "second effect" of pitch shift reported by earlier investigators. This hypothesis was verified by observing the dependence of the pitch shift on SPL and its reduction when masking noise was introduced in the region of the combination tones. Combination tones were thenceforth considered to be an important factor in the determination of the low pitch of complex tones. As paraphrased by Plomp (1976, p.124), Smoorenburg's data indicated that " combination tones, introduced by the ear, contribute to the low pitch of harmonic and inharmonic tone complexes just as though the stimulus included these frequencies".

2.8.10 Central origin of the pitch of complex tones: Evidence from dichotic listening

The basic foundation for the original "residue" hypothesis was attributed to temporal processing by the auditory system. The interaction of poorly-resolved high components of a complex sound produced

modulations of the waveform that corresponded to the periodicity. The following of this "beating" pattern was assumed to be the determinant of the perceived pitch.

The experimental results of Plomp (1967), Ritsma (1967) and Smoorenburg (1970) showed, however, that the pitch of complex tones could not be based on a strict beating hypothesis. Rather, a spectral basis was indicated, given the greater weighting of **resolved** components in contributing to the overall pitch.

An extreme example of resolution of components that prohibits any beating-type interaction is the case where components are presented **separately to the two ears.** This type of "dichotic" presentation was shown by Houtsma and Goldstein (1972) to be equally efficacious as monotic presentation in evoking a low-pitch percept.

Their experiments were designed to investigate the ability of the human auditory system to follow melodies based on relations between missing fundamentals of "residue"-type sounds presented sequentially. The complex tones used comprised two successive, upper harmonics (m, m+1) of the missing F0. Performance on a musical-interval identification task was seen to be essentially on par, for both monotic and dichotic presentation and was better when low harmonic numbers were used.

A simulation of the "inharmonic residue" experiment, using interval identification, rather than pitch matching as an indicator of perceived pitch, also verified the "first effect" of pitch shift for both monotic and dichotic presentation. The "second effect" however, was found

only for monotic and diotic conditions and thus appeared to reflect the influence of aural combination tones arising from within-channel interactions.

The fact that performance on the interval-identification task was basically similar for both the monotic and dichotic case and improved for lower harmonic numbers, lends further credence to the hypothesis that "channel separation" of components gives better estimates of the F0. This manifested importance of **frequency resolution as an essential part of the pitch extraction process** established beyond doubt that the "residue" hypothesis of Schouten (1939) had to be modified radically.

The dichotic results were also taken to imply that pitch was estimated by "a **central mechanism** that integrates and processes information from both cochleas" (p. 522). The central mechanism was considered to operate on neural signals derived from components resolved by the peripheral organs. While the exact nature of the neural mechanism mediating pitch extraction was not specified by Houtsma and Goldstein, it was believed to be consistent with both "place" and "time" coding schemes.

Under the former scheme, information about the frequencies of input signals would be preserved in the form of the place of active nerve fibers, while under the latter scheme, it would be preserved in the form of the temporal firing patterns of individual fibers (at least for lower frequencies) (Whitfield, 1970)).

2.8.11 Theories of pitch perception

The rich store of information obtained from the experiments conducted over the three decades following Schouten's "residue" hypothesis, provided a broad database with which to construct theories of pitch perception. Three theories that emerged contemporaneously are described below. While differing in details of implementation, these models are similar in spirit in that they treat the extraction of pitch as a pattern-recognition process. Each one acknowledges a two-stage process: a peripheral stage characterized by a limited frequency analysis, and a central stage in which the overall pitch is derived. While some details of the peripheral stage are provided, the nature of the latter "central" process is largely speculative. All three theories provide a reasonable fit to the data to different extents.

2.8.11.1 Goldstein's "optimum processor theory for the central formation of the pitch of complex tones"

Following the example of Siebert (1970), Goldstein (1973) applied principles of statistical estimation theory to model the behavioral and physiological data obtained from studies of pitch perception. The underlying basis for the development of his theory was the hypothesis that "the pitch of complex tones is necessarily mediated by a central processor that operates effectively *only* with signals derived from aurally

resolved simple tones".

The theory views a "central" pitch processor to be a "recognizer of spectral patterns supplied by the peripheral frequency analyzers". "The peripheral frequency analyzer extracts from complex-tone stimuli all constituent simple tones that differ in frequency from their neighbors by more than some resolution limit". The central processor then operates on the frequencies of the resolved harmonics using a "template" matching procedure.

Assuming periodic stimuli with adjacent harmonics, the pitch processor computes the harmonic numbers of the components present in the stimulus (inclusive of combination tones). A template of a form akin to a harmonic "comb" is then applied to the harmonic data to obtain the "optimal estimate" of the F0. This corresponds in essence to the gap between the "teeth" of the best fitting comb and is formally derived using a maximum- likelihood statistical estimation procedure.

In its original formulation, the theory assumed periodicity and the successivity of harmonics. The latter assumption was later extended to include non-successive harmonics (Gerson and Goldstein, 1978). In addition to these modified assumptions, a basic constraint of the theory is that "the estimator receives noisy information on the frequencies, but not amplitudes and phases, of aurally resolvable simple tones from the stimulus and its aural combination tones" (p. 1496).

A measure of this "noise" in estimating component frequencies is quantified by the parameter σ_k which is the standard deviation of the

mean frequency f_k obtained from the distribution of measures of the frequency of the k^{th} component. The parameter σ_k is the only free parameter of the model and is considered to be a function of a single variable, the component frequency f_k.

These theoretical assumptions were found to be good approximations in explaining a variety of empirical phenomena, ranging from the phenomenon of the "dominant region" to the precision in estimating the periodicity pitch of complex tones and the low pitch of inharmonic complexes.

The theory further showed that the precision of the estimate of the fundamental frequency of a complex is related to the precision in estimation of its' component frequencies f_k via the relation:

$$(fo/\sigma_0)^2 = \sum_{K=1}^{N} (f_k/\sigma_k)^2 \quad \text{[Goldstein, 1973; Eq. 15]}$$

The probability density function giving the estimate of the F0 (or "fo") is found to be crucially dependent on the probability of correct estimation of the harmonic numbers comprising a complex. High probability of correctly estimating the harmonic numbers in the stimulus leads to a precise measure of the stimulus period.

Both theoretical and empirical observations indicate that larger errors occur for estimation of the F0 with increasing harmonic numbers in the stimulus. These errors appear to be errors in judgment of the correct

harmonic "number". Thus for example, in the experiments of Schouten et al. (1962) where subjects were required to match the pitch of an inharmonic stimulus with equal spacing with that of a harmonic residue, the matching errors indicated that the listeners had started attending to a different order of the harmonic number "n" than the original number of the center component that was being shifted in frequency.

The relations between the probability of correctly estimating harmonic numbers, the order of harmonic numbers 'n', and the standard deviation σ are shown in the two representations shown in **figure 2.9.**

The relative standard deviation σ/f is the parameter in the upper frame, while lower harmonic number n is the parameter in the lower frame (from Goldstein, 1973). As can be seen, the probability of correct identification of harmonic numbers is greater for lower harmonics where cochlear spectral resolution is greatest.

2.8.11.2 The pattern transformation model of pitch

The importance of lower harmonic numbers in contributing to the pitch of a complex is also manifested in a number of relations predicted by the "pattern transformation" model of Wightman (1973).
This spectrally-based model describes the process of pitch perception in terms of a sequence of transformations of "patterns of neural activity".

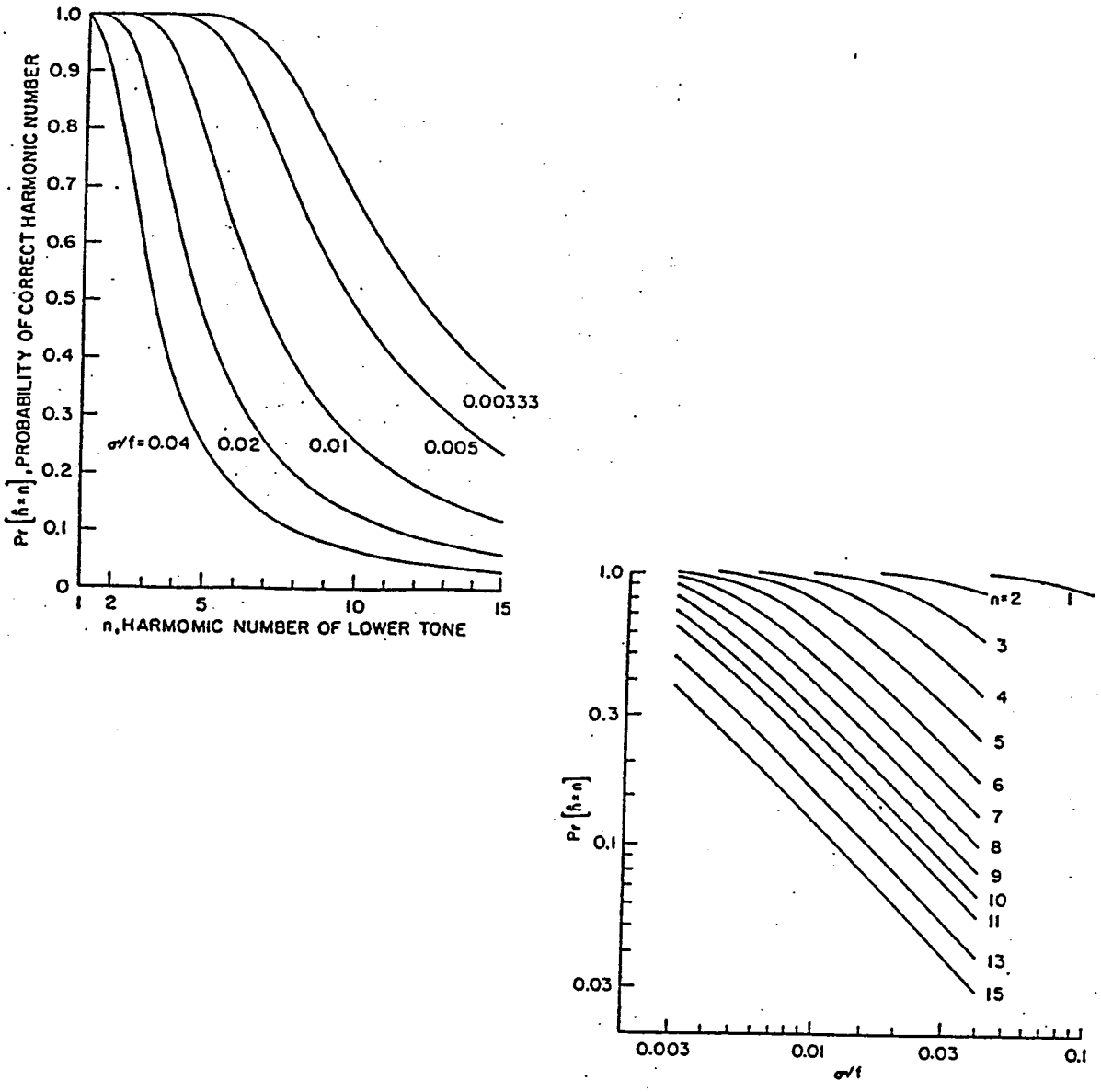


Figure 2.9 Probability of correct estimation of harmonic numbers in a two-tone signal with frequencies nF0 and (n+1)F0. In the top frame, the parameter is the relative standard deviation of the random signals representing the component frequencies. In the bottom frame, the parameter is the lower harmonic number. Both representations indicate a higher probability of correct identification of lower harmonic numbers. [From Goldstein, 1973]

The term "pattern" is used to refer to a hypothetical 2-dimensional distribution of neural activity in some ensemble of nerves, where the two dimensions reflect "place" and "amount" of activity.

The "type" of neural activity at each place (such as the temporal distribution of nerve firings) is ignored. The peripheral pattern is assumed to roughly represent the power spectrum of the stimulus and is derived from it by convolution with a triangular, constant Q filter (to simulate critical band-type filtering). The peripheral pattern is then assumed to be Fourier-transformed into another pattern of activity that roughly represents the autocorrelation function of the stimulus. Pitch is derived from the positions of maximal activity in this transformed pattern. The model is phase-insensitive and the temporal fine structure of the stimulus is virtually ignored.

The model is useful in predicting situations where the pitch evoked will be ambiguous (as for inharmonic stimuli), or conversely, where it will be strong (as for harmonic stimuli). "Pitch strength" is related to maxima in the transformed pattern. The frequency corresponding to the maximum is a measure of the "pitch value" assigned to the stimulus. Better resolution of components is considered to contribute to a "stronger" pitch, while pitch strength is predicted to decline as the stimulus becomes higher in overall frequency and as the number of components in the stimulus is decreased.

2.8.11.3 Terhardt's theory of "virtual" and "spectral" pitch

In describing pitch phenomena associated with complex tones, Terhardt (1974) deemed it necessary to distinguish between **two kinds of pitch**: spectral pitch and virtual pitch. Spectral pitch is the pitch associated with a pure tone (or a partial), while virtual pitch is the overall pitch associated with a complex tone.

In line with von Helmholtz, Terhardt also distinguishes between **two kinds of modes of pitch perception**: the analytic (resulting in perception of spectral pitch), and the synthetic (resulting in perception of virtual pitch). Both these types of pitch are considered to be derived from spectral cues, but according to different derivational principles. "Virtual" pitch is considered an attribute which is the product of auditory ***gestalt*** perception. The principle of gestalt perception is carried over to the field of audition by Terhardt, based on the analogy that a ("virtual") tone may be perceived in some instances, just as a visual "contour' may be perceived even when not actually present. Perception of "residue" pitch is a manifestation of such an auditory gestalt.

A "learning matrix" constitutes the core of Terhardt's model. Repeated exposure to harmonic sounds (such as voiced speech sounds) and their associated spectral pitch cues impresses "traces" upon the matrix. The nominal spectral pitches of the components are numerically equal to their signal frequencies, while the "true" spectral pitches differ

from the nominal ones by some "pitch shift" factor. "Virtual" pitch is evoked when these traces are activated by an incoming group of spectral cues. Each spectral-pitch cue produces eight virtual-pitch cues. Alignment of these cues at a particular value of virtual pitch yields a "maximum" in the virtual-pitch function which is then assigned to be the virtual pitch H (in pitch units "pu").

This esoteric scheme can be considered in more simplistic terms to be equivalent to a "sub-harmonic" matching procedure. The sub-harmonic which is the highest common factor of the components of a complex is taken to be the "virtual" pitch. Algorithms that provide detailed guidelines for calculating virtual pitch of complex stimuli taking into account factors such as sound pressure level and inter-component masking have been provided by Terhardt (1979) and Terhardt et al. (1982 a, b). The latter of these has been extended to address issues such as:

1). Extraction of spectral data (frequencies and amplitudes) from an input signal.
2). The relative salience of competing virtual pitches.
3). The contribution of factors such as harmonic number, harmonicity, aural resolution and spectral dominance of components to the "salience" or relative "weight" of a candidate virtual pitch.
4). The role of spectral pitches as not just contributors, but **competitors** of virtual pitch.

The explicit acknowledgment of this last issue is (in this author's opinion) the greatest contribution of the "virtual pitch" theory. Terhardt's model accepts at the very outset that the "analytic" and synthetic, or "holisitic" perceptual modes may compete. Neither mode will be completely suppressed in an actual listening situation. The "whole pitch percept is described as a competition between spectral and virtual pitches" (Terhardt et al. 1982, p. 686).

2.8.12 Recapitulation: Pitch and frequency

As indicated in the preceding review, there have been periodic(!) changes in explanations for the perception of the "low" pitch of complex tones throughout the last century. The supremacy of the fundamental component as the fundamental bearer of pitch has been disputed by the observation that the pitch of complex tones is not always trivially related to the F0. The existence of a low pitch for both harmonic complexes lacking the fundamental component, and for inharmonic complexes with no common denominator indicate that both "place" and "period" cues are used in the estimation of pitch.

Schouten's residue hypothesis attributed the perception of low pitch to the *failure* of the auditory system to resolve the higher harmonics of a complex. However, the experiments of Houtsma and Goldstein (1972) forced the abandonment of this explanation in favor of better-substantiated theories that demonstrated that, on the contrary, it

is the **resolved**, rather than the unresolved components that dominate the sensation of pitch. As stated by Plomp (1976), "it has become more and more clear, that the ear's frequency resolution has to be considered as **contributary rather than a disruptive** factor in pitch extraction" (*sic.* p. 111).

The pitch theories reviewed above, indicate that **pitch, like timbre, is derived from the spectrum.** Ironically, these theories ignore the possibility that these perceptual features of complex sounds may interact, given their common spectral basis. The issue of timbre and pitch interactions is addressed in the following section.

2.9 Interaction of timbre and pitch

In several of the experiments on pitch perception reviewed in the last section, changes in timbre probably ensued on many occasions when frequency changes were made in the stimuli used. Since timbre was not the focus of those pitch studies, this effect was typically ignored or referred to only anecdotally. Similarly, in many studies of timbre, pitch was restricted to certain defined values (such as F0=200 Hz used by von Bismarck (1974), and E^b used by Grey (1975)).

The official definition of timbre, as recorded by the American Standards Association (ASA, 1960) states it to be: **"that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness**

and pitch are dissimilar". Using a single pitch value in studies of timbre, satisfies the ASA definition separating timbre from other percepts such as pitch and loudness, but results in loss of information about possible relations between pitch and timbre.

Some researchers have shown pitch and timbre to be independent for particular sets of stimuli (Miller and Carterette, 1975; Plomp and Steeneken, 1971). However, these percepts have also been observed to covary for many natural instruments, for example, those in which the spread and shape of the spectral distribution changes with fundamental frequency of excitation (Benade, 1986; Risset and Wessel, 1982; Wessel et al. 1987).

A note appended to the negatively-phrased definition of timbre (ASA, 1960, p.45) captures the essence of the possible confounding of pitch and timbre. The note acknowledges that: "Timbre depends primarily upon the spectrum of the stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum and the temporal characteristics of the stimulus".

This extended definition of timbre may equally well be applied to the pitch of complex tones. Pitch too is derived from the spectrum, is correlated with the periodicity of the waveform, and is affected by sound pressure, frequency location of the spectrum, and temporal characteristics of the stimulus !

It thus appears that these oft-quoted perceptual "dimensions" are in fact related, and may well interact with each other when changes are

made in associated physical dimensions.

2.9.1 Influence of timbre on pitch matching

Many experimenters studying pitch mention that subjects have difficulties making pitch matches when the spectra of the test stimulus and the reference stimulus vary greatly, as in the case of a complex tone being matched in pitch with a sinusoid. Davis et al. (1951) reported that all their listeners trying to match the low (≈130 Hz), repetition frequency of 2000-Hz tone pips to that of a pure tone commented that "matching is difficult because the sounds are so unlike", an observation that has been reported by several other researchers as well.

Anybody who has tried to match pitches of sounds with different perceived timbres, knows intuitively that it is easier to match pitch within the context of a similar timbre than it is across timbres. The latter task is not impossible, considering that members of musical performance groups do routinely make such pitch matches across different instruments in the act of "tuning up". It is simply perceived as being more difficult.

This commonplace experience was verified experimentally by Crowder (1989). In a study on "imagery" for musical timbre, he showed that "people more easily judge that two tones have the same pitch when they are also of the same instrumental timbre, than when not". In that study, listeners were required to judge if the pitch of two sounds

presented sequentially was same or different. The sounds were actual samples of musical notes recorded from 3 different instruments (guitar, flute and trumpet). Reaction time was used as an indicator of ease of judgment. It was observed that listeners were quicker and more accurate in judging that the musical "pitch" of the two notes was the same, when they were derived from the same instrumental timbre.

Crowder also reported that three of the twelve listeners used apparently had trouble performing the task. They were "reluctant or unable to identify two pitches as the same", when articulated by different instruments. Conversely, they had low accuracy scores in judging that the two notes were of different "pitch" when articulated by the same instrument. It seems that these listeners (dismissed as "outliers" by Crowder) were actually doing a **timbre** discrimination task rather than the "pitch" discrimination task they had been assigned. This is further corroborated by the improved accuracy and speed of their judgments when the two tones were derived from the same instrument playing the same note, or a different instrument playing a different note. In these case, there was no conflict between timbre difference and pitch difference and their improved performance was on par with other listeners.

2.9.2 Influence of pitch on timbre identification

The studies of Saldanha and Corso (1964) and Clark et al. (1964)

reviewed in the earlier sections on timbre revealed some interesting auxilliary facts about the relation between identifiability of instrumental timbre and the pitch being articulated.

The pitch of recorded tones (related to the F0) affected timbre identification in the experiment of Saldanha and Corso. More correct judgments were obtained at F0 corresponding to F4 (≈349 Hz) than for F0 corresponding to C4 (≈262 Hz) or A4 (≈440 Hz). In the light of the formant theory described in section 2.4.1, one can hypothesize that some values of the F0 (related to spectral density of components) could have led to a better "tracing" of the spectral envelope because of components lining up with formant frequencies

Similarly, Clark et al. observed a tendency for all instruments to be identified as being identical at high notes. Since the spectrum becomes more sparse with wider component spacing at high values of F0, fewer cues may be available to define the shape of the spectral envelope and differently-shaped envelopes may thus be perceived as being the same because of the lack of definition of the spectral form.

2.9.3 Spectral pitch, or timbre ?

Many of the experiments on pitch perception reviewed above, were motivated by curiosity about the general functioning of the hearing mechanism and its ability to encode stimulus features such as frequency and intensity. Pitch happens to be a percept that lends itself well to

quantifiable comparisons and thus proved to be a useful tool with which to probe these processes.

Most experimenters used pitch comparisons as criteria for performance, assuming that the concept of pitch was intuitively understood by listeners. Indeed, in most cases pitch does emerge as a clear, well-defined aspect of a periodic stimulus. However, the conflict between "place" and "periodicity" pitch and the ambiguous, deviant pitch observed for inharmonic stimuli highlight the fact that there are some **confusing aspects** of pitch perception. This observed confusion raises questions about "what" exactly was being compared by listeners in some of the earlier studies: **was it pitch, or was it something else ?**

The fact that there are **two aspects of pitch** has been acknowledged in the field of pitch research (see section 2.8.5). Thus, "virtual" pitch or "periodicity" pitch is one aspect of pitch sensation that usually corresponds to the overall "musical" pitch of a complex. Another aspect of pitch is "spectral" pitch which corresponds to the frequency of components comprising a complex.

To avoid the disturbing influence of differences in "quality" or "timbre", on matching pitches of stimuli with vastly differing spectral composition (as for a pure tone and a complex tone), some experimenters chose to use complex tones as matching signals. However, in that case, another problem was manifested; some listeners perceived the spectral pitches of components rather than the virtual pitch of the complex as a whole and performed the task on the basis of spectral pitch matches. The

results of Schouten et al. (1962), Smoorenburg (1970), Risset (1971, 1986) and Gerson and Goldstein (1978) show direct evidence of such confusions.

The circular irony of this dilemma in using appropriate reference signals for pitch comparisons seems to have eluded the investigators. The original, standard scheme for obtaining pitch matches (as advised in the ASA (1960) standard) was to match the pitch of a complex to that of a sinusoidal signal with unambiguous frequency. It was the difference in timbre between such stimuli that prompted the use of complex matching signals. But the eventual choice of complex signals that lacked corresponding components again resulted in timbre differences between test and comparison signals, thus leading back to the original dilemma !

At the heart of this confusion between spectral pitch matching versus virtual pitch matching, and the influence of timbre differences, is the fact that **the spectrum of a sound is directly related to both *spectral* pitch and timbre.** Confusions between spectral pitch and timbre may thus arise because of their common physical basis. Further, spectral pitch contributes to virtual pitch. Thus interactions between the timbre and the overall pitch of a sound may also be observed. Timbre and pitch may therefore not be strictly independent features of sounds.

The originator of the "residue" hypothesis - J. F. Schouten himself, in a paper with colleagues Ritsma and Cardozo (1962, p.1419), mentioned that:

"Hitherto, the concepts "timbre" and "pitch" have been used

rather vaguely. In fact, many a subject when first listening to a residue is at a loss whether to describe the sound as "high" or "low", since it seems to carry both aspects simultaneously. Consider, for example, a periodic impulse with a repetition frequency of 200 cps. The Fourier components have frequencies 200, 400, 600, ... cps. If a filter is applied, passing only the components 1200, 1400 and 1600 cps, a residue is heard with a pitch equal to that of the fundamental. Again, if the filter passes the components 1800, 2000 and 2200 cps, a residue of equal pitch is heard. Both residues are easily distinguishable, the latter having a sharper timbre than the former.

... In subjective sound analysis, the perception of the residue often predominates over that of the lower harmonics and determines the timbre as well as the pitch of the sound".

Having made these insightful statements about the common underlying basis for timbre and residue pitch, Schouten et al. went on to separate these percepts by using an operational definition relegating pitch to be (missing) F0 and timbre to be associated with the physical components actually present in the spectrum , and timbral consequences of frequency changes were not discussed in the later exposition of the experiments.

The relation of timbre to spectral pitch was highlighted in an interesting experiment by Smoorenburg (1970)designed to study the relative discriminability of "residue" pitch and the place pitch

corresponding to frequency of components. He presented listeners with a pair of 2-frequency complexes with f1 and f2 for two tones as given below:

Tone 1	Tone 2
f1 = 1800 Hz	f1 = 1750 Hz
f2 = 2000 Hz	f2 = 2000 Hz

The subjects were asked to report if the pitch of the second sound was higher or lower than that of the first.

The rationale behind the design of this interesting stimulus was the fact that two different responses were possible, depending on whether listeners were judging relations between frequencies of components, or the pitches of the complexes as a whole. Thus, if a listener reported the second sound as being higher in pitch, then the judgment would indicate a comparison of residue pitch (F0 = 200 Hz vs. 250 Hz). However, if the listener judged the first sound to be higher, then a comparison of component frequencies (1800 Hz vs. 1750 Hz) would be implicated.

Smoorenburg found substantial individual differences across listeners. Results indicated that almost equal numbers of listeners reported each type of pitch change ! Some followed the change in component frequency as a cue for pitch change, while others followed the "true" F0 change of the complex as a whole. Listeners thus varied greatly

in making judgments about pitch change.

Smoorenburg inferred that "the perception of a pitch jump corresponding to the fundamental frequencies was based upon pitches of the complex tones perceived as a whole and the judgments in opposite direction were based upon the pitches of individual part-tones or perhaps **just upon timbre**" !

This marginal allusion to "timbre" comparison as a possible basis for the judgments obtained is **a dangerous indication** that in many of the experiments that purported to be studying "pitch", it may in fact have been "timbre" that was being judged by listeners.

Another example of spectral pitch matching and the confounding of pitch and timbre was provided by Risset (1971, 1986). He presented listeners with a two-tone sequence, with frequencies of components for the tones as given below (in Hz; taken from Risset, 1986):

Tone A'	49.6	102.4	211.2	435.2	896.0	1843.2	3788.8
Tone B'	99.2	204.8	422.4	870.4	1792.0	3686.4	7577.6

The frequencies of components of tone B' were double the value of those in tone A'. Despite this octave change upwards, however, **50 out of 53** listeners reported that tone B' was *lower* in pitch than tone A' ! Some listeners also commented that "B' was more "brilliant", or "***higher in timbre***" (Risset, 1986, p. 961; emphasis, mine).

It appears that the listeners in Risset's experiment were also making spectral pitch comparisons as did some listeners in the other studies mentioned above. Instead of comparing the "corresponding" harmonics "n" of the two tones A' and B', the pitches of components corresponding to the $n+1^{th}$ components of A' appear to have been compared with the n^{th} components of B'. B' was thus judged to be lower in pitch.

However, B' occupied a higher spectral locus than A', and also contained a higher harmonic unmatched with the components of A'. This difference in absolute frequency may have led to a perceived change in a timbral attribute such as "sharpness" that is directly related to the upper frequency limit of the spectral locus (von Bismarck, 1974 b; see section 2.7.5.1).

Listeners' use of the adjective "higher" to describe a timbre difference, is indicative of a perceived change in a *pitch-like aspect of timbre* (such as "sharpness"), that can also be scaled perceptually along an ordinal scale from low to high. Additional observations with stimuli of this type led Risset (1978, p.18) to state quite boldly, that **timbre is in fact just another "name" for spectral pitch . . .**

2.10 Discrimination of frequency changes in complex sounds

Given the dependence of both pitch and timbre on the spectrum, it

would be expected that **changes** in the frequencies of spectral components may be construed as changes in pitch, in timbre, or in both these percepts simultaneously.

The ability to discriminate frequency changes has been the focus of extensive research in psychoacoustics. Evans and Wilson (1974, p. 173) contrast the synonymous terms "frequency selectivity", "frequency analysis" and "frequency resolution", which imply the ability to separate or *resolve* simultaneously presented components of a complex sound with **"frequency discrimination"**, which is defined as the ability to detect differences in frequency between signals presented **non-simultaneously.**

The former abilities are usually cited as being the major bases for detection of signals in noise (Moore, 1982), perception of pitch (Goldstein, 1973), perception of "roughness" (Terhardt, 1974), "brightness" (Lichte, 1941), perception of consonance and dissonance (Plomp and Levelt, 1965), and other perceptual phenomena incorporating sound events presented simultaneously. Frequency discrimination on the other hand, generally implies a comparison across *successive* sound events.

The ability to distinguish between sounds based on sequential frequency changes comes into play in a variety of perceptual tasks such as pitch discrimination (Harris, 1952) timbre discrimination (Grey, 1978), pattern discrimination (Watson et al., 1975, 1976), multiple- interval signal detection tasks (Elliott, 1962), and a variety of other phenomena, including rhythmic accenting in sequences based on pitch or timbre

changes (Garner and Gottwald, 1968; Handel, 1974).

As was mentioned in section 1.2 and 1.3, frequency changes in complex sounds can be brought about in many ways, and unlike the case for pure-tone stimuli, may not always evoke changes in pitch. The perceptual correlates of frequency changes that enable discrimination are hard to ascertain and have not been the major focus of studies primarily aimed at finding "just noticeable differences".

The lack of a viable vocabulary to describe these correlates is a major handicap for studies employing complex signals that lack linguistic meaning. For speechlike sounds however, the availability of phonetic labels can be exploited to describe perceptual changes accompanying stimulus changes. Thus, the F0 of a vowel sound, or the resonance frequency corresponding to one or more of its formants can be changed and the effect on pitch, quality, or phonetic identity of the vowel can be studied.

2.10.1 Frequency discrimination and speech perception

The technological advent of telecommunication systems some forty years ago, imposed some practical demands on the study of speech sound discrimination. In order to determine the precision requirements for speech transmission in such systems, information about constraints of human auditory processing was needed. In this endeavour, a number of studies emerged, including some on **pitch discrimination** of vowels

(Flanagan and Saslow, 1958), and **difference limens for vowel formant frequency** (Stevens, 1952, Flanagan, 1955).

Flanagan (1955) synthesized vowel-like stimuli comprising four formants. Listeners were required to discriminate between a standard synthetic vowel and a complex containing a formant displaced from its standard frequency, by judging them to be "same" or "different". Based on these quality judgements, Flanagan found the DL for the first two formants (F1 and F2) to be approximately **3-5%** of the formant frequency.

In another experiment, Flanagan and Saslow (1958) employed similar four-formant synthetic vowel stimuli, but sought instead to determine the DL for fundamental frequency (F0), rather than formant frequency. The task required listeners to identify the pitch of the second tone of a pair of vowel-like tones presented sequentially as being "higher" or "lower" than the pitch of the first tone. Thus, explicit judgements of *pitch* were asked for, even for cases where the F0 of the two tones was the same! The DL for F0 discrimination under this task was found to be **0.3-0.5%** , an order of magnitude *smaller* (i.e. more acute) than the DL for formant frequency.

Flanagan and Saslow also used a pure tone signal with F0 equivalent to that of one of their complexes (120 Hz). They found the DL for the pure tone to be slightly larger than that for the F0 of the vowel-like stimuli. The better performance observed for the complex stimuli led them to suggest that listeners make some use of frequency

changes in higher harmonics when discriminating the pitch of vowels, and do not depend solely on changes in frequency of the fundamental component. Listeners are apparently able to listen at the frequency region that would yield optimal cues for discrimination.

Scheffers (1983, p.35) summarizes some of these issues in the statement that listeners can "in principle use three cues to discriminate differences in the fundamental frequency of vowel sounds, viz. a difference in the residue pitch of the sound, a difference in the pitch of a single harmonic, or a shift in the spectral envelope". (The term "residue" is used in a historic sense, to indicate the overall "low" pitch or "periodicity" pitch of the sound, rather than the pitch resulting from combined interaction of unresolved high harmonics as Schouten (1939) had originally proposed).

It thus appears that in discriminating complex tones (such as vowels), a listener has many potential cues available to aid the task. The multiplicity and salience of these cues will depend on the particular spectral composition of the complex, and the manner in which it is changed. Mermelstein (1978) reported such a **context effect** in discrimination of formant frequency for steady-state vowels and vowels bound in a consonantal frame. The DLs for formants of time-varying vowels embedded in a consonant-vowel-consonant (CVC) context were found to be significantly larger than those for steady-state vowels. Further, the increment in the DL varied, dependent on the particular consonant used for the context. Mermelstein suggests that "the

difference in DL values in and out of context has, at least partially, an auditory origin".

Gagne and Zurek (1988) further pursued this idea of resonance-frequency discrimination reflecting an aspect of general, auditory discrimination, rather than being a limited, phonetically- based effect. They reported measurements of the jnd in resonance frequency (ΔFr) for a second-order filter excited by a source signal that was either periodic with a fixed F0, periodic with a smoothly gliding F0 (over ΔF0=50 Hz), or aperiodic, random, white noise. While the different source waveforms produced roughly the same discrimination performance (summarized by $\Delta Fr = .079\ Fr/\ Q$, where Q = 1.0 to 36.0), different perceptual cues seem to have been operative for the different waveforms.

For the fixed, periodic source, the change in Fr was signalled by relative amplitude changes of the spectral components. Perceptually, this change seems to have been construed as a change in "quality" or timbre. For a continuous-spectrum source, as filter bandwidth is narrowed, the perceptual distinction apparently changed from being one based on a timbre change to being one based on a change in pitch.

It was thus concluded that identifiable perceptual cues could indeed be used in discriminating resonance frequency. It was also surmised that a *combination* of possible timbre- and pitch-like percepts could guide the discrimination of signals in which both F0 and Fr were changed simultaneously.

2.10.2 Place-periodicity dilemma revisited

The data of Flanagan and Saslow (1958) showed discrimination of F0 for vowels to be finer than that for formant frequency discrimination or for discrimination of a pure-tone of equivalent F0, by an order of magnitude (0.3-0.5% instead of 3-5%). This discrepancy has often been generalized to imply that **pitch discrimination is better for complex tones than for pure tones.**

Ritsma (1963) however, warned against such generalization, arguing that the data of Flanagan and Saslow did not give conclusive evidence about the accuracy of **"periodicity pitch"** discrimination. Rather, he claimed that "their results can be more readily interpreted by assuming that the subject, in determining the jnd in pitch of a vowel reacts to a shift of those components at which the jnd in pitch is smallest, rather than to a change of periodicity pitch" (p. 34).

A change in F0 of a harmonic complex (such as a synthetic vowel) results in simultaneous changes in frequency of all the harmonics. It is therefore not clear if listeners in the "pitch discrimination" experiment of Flanagan and Saslow made judgements based on a comparison of the overall pitch of the sound (related to F0), or on the basis of multiple comparisons between spectral pitches corresponding to component frequencies.

Flanagan and Saslow themselves conceded that the "slightly more

acute discrimination of a change in the F0 of vowels than in a pure tone of equivalent frequency and level", probably resulted from "advantages accruing to discrimination from the relatively large changes in harmonic frequency". These large changes come about because of the *proportional* variation of harmonic frequencies with changes in F0.

The frequency of a harmonic 'n' is $f_n = n \times F0$. If F0 changes by an amount = Δf, the frequency of the harmonic will correspondingly change by an amount equal to n times the change in F0:

$$\Delta f_n = [n \times (F0 \pm \Delta F0) - (n \times F0)] = \pm (n \times \Delta F0).$$

The first harmonic (i.e. the fundamental component), thus undergoes the smallest change in terms of *absolute* frequency Δf, while higher components undergo larger changes. Flanagan and Saslow commented that phenomena such as inter-component masking and variation in amplitudes of components due to changes in formant frequencies could lead to degradation of the higher harmonic "advantages". Otherwise, the DL for "vowel pitch" should have been *even smaller* than that obtained in their study.

The existence of harmonic advantages for stimuli employed in a F0 discrimination task presents the familiar dilemma noted in section 2.9: ***Was periodicity pitch being compared, or was the place pitch corresponding to individual components being compared?***

2.10.3 Pitch discrimination of residue tones

Place and periodicity cues were confounded for the vowel sounds used by Flanagan and Saslow. Changes in "place" occurred concurrently with changes in periodicity. For **residue** stimuli, these factors can be changed independently, to some extent, allowing for examination of their relative contributions to discrimination judgments.

Thus, the frequencies of spectral components of a harmonic complex can be changed, while keeping the same missing F0. Or the missing F0 can be changed while keeping *roughly* the same spectral locus. Since F0 change is usually equated to pitch change for harmonic sounds, a change in missing F0 may yield a change in the "residue pitch" associated with the stimulus.

Ritsma (1963) used residue tones as stimuli in a pitch discrimination paradigm, to enable judgments about accuracy of "periodicity pitch" without interfering place cues. Since the tones could be made to have the same (missing) F0, while comprising components in different spectral regions, it was assumed that comparisons of frequencies of corresponding components would be prevented, and listeners would be forced to rely on comparison of F0.

Ritsma's listeners were required to match the pitch of two such residue tones given an initial difference in the underlying F0 and a difference in the *harmonic numbers* present in the tones. Comparison of place pitch was considered to be prevented by elimination of "common"

components between test tones. Thus, for example, the two tones being compared were typically assigned harmonic numbers 6,7,8 and 9,10,11 respectively. Contrary to the observations of Flanagan and Saslow, pitch discrimination for such complex tones was found to be much **poorer** than that for pure tones of equivalent F0.

Ritsma's results played an important role in the development of the "optimum processor" model of pitch perception (Goldstein, 1973; see section 2.8.11.1 for details). The higher DLs obtained for residue tones conflicted with the prediction made under the model that precision of the estimate of fundamental frequency F0 should be better than the relative precision of the estimate of the frequency of any component in the complex (p.1501, Eq.15: $(fo/\sigma_o)^2 = \sum_{k=1}^{N} (f_k/\sigma_k)^2$).

Goldstein interpreted Ritsma's data as demonstrating that "the precision for periodicity pitch is not equal to, nor theoretically accounted for by the precision with which frequency of simple tones can be discriminated" (p.1506). Using the additional data of Houtsma and Goldstein (1972), Goldstein (1973) thus concluded that the variance functions describing the precision with which frequency information is conveyed to the central processor for periodicity pitch, are much greater (about 5 times) than those measured for simple tone frequency discrimination. This greater variance is also in accord with the data of Ritsma (1963).

The accuracy of the pitch estimation process is also impaired with increasing harmonic number n and increased spectral spacing of

components (Gerson and Goldstein, 1978; Goldstein, 1973). Residue tones comprising high values of n would thus have less salient pitches. Discrimination of residue pitch should thus **get worse** as harmonics comprising the tone get higher. This has indeed been reported by Hoekstra 1973, (cited by Nordmark, 1978, p. 268). The DL for missing F0 was reported to increase as periodic pulses were selectively filtered to contain higher harmonics ($\geq$ n=8).

2.10.4 Spectral pitch -- timbre -- virtual pitch conflict revisited

Moore et al. (1984) questioned Goldstein's assumption of "noisy" channels conveying information from the periphery to the central processor based on the data of Ritsma (1963). Claiming that the "**marked difference in timbre**" between the residue tones used in Ritsma's experiment detracted from the underlying pitch difference, they cited a number of studies showing finer discrimination for complex tones than for pure tones of equivalent F0 (e.g. Henning and Grosberg, 1968; Fastl and Weinberger, 1981).

These studies also indicated that higher harmonics mediate discrimination for complex tones with low fundamental frequency (for F0$\leq$2000 Hz). The experiment of Henning and Grosberg in particular, suggested that the periodicity of a complex tone is discriminated on the basis of its single most discriminable harmonic. The discriminability of a

"harmonic" however, was deduced indirectly, from measurement of the DL for a pure tone of frequency equal to that of the harmonic.

It has been pointed out by Terhardt (1979), that the components within a complex signal can mask each other. Moore et al. (1984) thus proposed that a more appropriate comparison to the DL for a complex tone would be the DL measured for components *within* the tone, rather than the DL for components presented in isolation. To this end, they conducted an experiment in which listeners were asked to indicate which tone out of a pair of tones presented sequentially was "higher" in pitch, based on changes in F0, or in the frequency of a single component.

Three "standard" complex tones comprising equal-amplitude harmonics of a 200 Hz fundamental were used. These differed in the number and region of components, containing either harmonics 1-7, 1-12, or 5-12. The DL was found to be lowest when changes were made in the frequency of harmonics lower than n=5. For higher harmonics (n= 5 to12) the DL increased abruptly, from less than 1% to around 5%. The highest harmonic however, was also well discriminated.

The DLs for complex tones as a whole, were generally found to be lower than the DL for their most discriminable components, suggesting that information must be combined across harmonics. Further, the DLs obtained for the complex tones as a whole, were slightly smaller than the DLs for individual components measured in this way. Based on these data, Moore et al. suggest that the relation derived between estimation of F0 and estimation of component frequencies f_k in Goldstein's model

(Eq.15) is valid as a predictor of the DL for periodicity pitch from component DLs, without having " to assume that the channels conveying information from the periphery to the central processor are noisy" (p. 560).

In an ironic replay of past events, Faulkner (1985) questioned the results of Moore et al. (1984) based on reasoning similar to that of Ritsma (1963). He alleged that in discriminating complex tones with "corresponding" harmonics (i.e. harmonics having the same numerical ratio with respect to F0), listeners actually compare the **pitches of individual components**, rather than residue pitches. In this case, the DLs for the complexes are found to be lower than the DLs for any single component frequency due to integration of information across multiple frequency bands. However, when the complex tones to be discriminated do not have corresponding harmonics, their discriminability is reduced.

The loss of acuity is ascribed to the lack of common or "coincident" components, that forces listeners to rely on the residue pitches for discrimination. Degradation in performance for this **true** residue pitch discrimination is consistent with, and reinforces the idea of an internal noise process as proposed by Goldstein (1973).

Faulkner's allegation has serious implicatons for research on pitch discrimination of complex tones with corresponding components, suggesting that the **pitch** of such signals is in fact not being compared at all !

Moore (1987), however, dismissed this suggestion, claiming that

the degraded performance for residue tones is due to marked **timbre differences** between such tones, that make pitch comparison difficult, rather than to the due to the absence of corresponding components per se.

Timbre is strongly related to the position and shape of the spectral envelope. A change in timbre may thus ensue if spectral frequencies of components are changed. Moore's criticism about the dissimilarity of timbre for residue sounds is thus a valid argument.

In fact, in the paper reporting "pitch discrimination of residue tones", Ritsma (1963) himself stated that "in the case of steady-state signals the place of maximal stimulation may be taken to be a measure of timbre whereas the pitch will depend on the quasi-periodicity of the signal" (p.34). For concurrent changes in both missing F0 and spectral locus, an interaction of timbre and pitch may thus indeed be taking place, but is not referred to at all in Ritsma's later exposition of the experiment.

The debate on relative discriminability of complex tones and their harmonics, started by Ritsma, and rekindled by Moore et al. and by Faulkner, still prevails. The fact that **both pitch and timbre may be affected by changes in F0 and in the frequencies of spectral components** again seems to be the factor responsible for this controversy. The role of corresponding components between tones compared for residue pitch, and the influence of timbre on such comparison needs to be explored to resolve the controversy. This was the aim of the first experiment reported in this dissertation.

2.10.5 Frequency discrimination and perceptual fusion

In the component-DL experiment of Moore et al. (1984), a paradigm was selected in which a **single** component was shifted in frequency against a backdrop of unchanged harmonics. In essence, the target component was "mistuned" from its harmonic frequency and the complex sound was thus rendered inharmonic. The only response option provided to the listener was to report which of two successive sounds "had the higher pitch". Feedback was provided for "correct" judgments.

This task and procedure seem very inappropriate, given the inharmonic nature of the stimulus used, and the fact that frequency changes are not always monotonically related to perceived changes in pitch (Risset, 1971).

As discussed in the section on perceptual fusion (2.4.1), inharmonic changes lead to perceptual segregation of a complex. Asking listeners to make judgments about pitch assumes a "synthetic" mode of listening. However, other discriminable changes in the stimulus may be available to the listener via an "analytic" listening mode (Houtsma, 1979). Such changes may be related to the perceived degree of fusion of components, or to timbral cues caused by "beating"- type interference of high harmonics. A listener would be able to discriminate complex sounds based on these **other perceptual changes**, rather than on the basis of perceived changes in pitch. Providing feedback for correct discrimination of a change in component frequency, may ironically reinforce the

listener's use of cues other than pitch. The experiment may then end up not having investigated "pitch" discrimination at all !

Moore et al.'s listeners did indeed report that changes in frequency of different harmonics were correlated with different perceptual cues. Lower harmonics were audible as separate tones when shifted in frequency, while the higher harmonics were not. The lower DLs obtained for lower harmonics may thus have been confounded with "fission" cues. The lack of such cues for the harder-to-"hear out" higher harmonics was similarly correlated with higher DLs.

2.10.6 Multiplicity of percepts associated with spectral changes in complex tones

Reports of percepts other than pitch associated with frequency changes in a complex sound, appear to have led Moore et al. (1985 a,b, 1986) to re-examine the stimuli and task used in their DL experiment (1984). Three related experiments followed, with different tasks assigned to the same basic stimulus set. In one case the relative dominance of components in contributing to the overall pitch was investigated in a pitch matching task (1985 a); in another, judgments of perceived "mistuning" or "detection of inharmonicity" were required in a two-interval-two-alternative-forced choice (2I2AFC) paradigm, and in a third case, thresholds for "hearing out" harmonics were determined in a single-interval paradigm.

Different results were obtained for discrimination of frequency change, dependent on the nature of the assigned task. For judgments of **perceived mistuning** for example, thresholds were seen to *decrease* with increasing harmonic number, in contrast to the DL experiment, where they increased.

The different results obtained for these two experiments are illustrated in **figure 2.10**. The top frame (taken from Moore et al., 1984) shows the frequency DLs (in %) as a function of the number of the harmonic that was mistuned. The bottom frame (from Moore et al., 1985 b) shows thresholds for detection of inharmonicity as a function of harmonic number.

The difference in the results of the two experiments appears to be directly related to the description of the assigned task.
Asking listeners to detect "inharmonicity" may set up the expectation of hearing "roughness" or "dissonance" to signal discrimination. Such cues are available for the poorly resolved higher harmonics, but not for the lower harmonics. Improvement in performance for the higher harmonics is thus observed. For the inappropriate "pitch" discrimination task assigned in the DL experiment, the better-resolved lower harmonics would be expected to be more likely to yield pitch cues. Accordingly, smaller DLs were obtained for the lower harmonics.

The relative dominance of low partials in determining the pitch of a complex tone was verified in one of the other experiments (Moore et al., 1985 a).

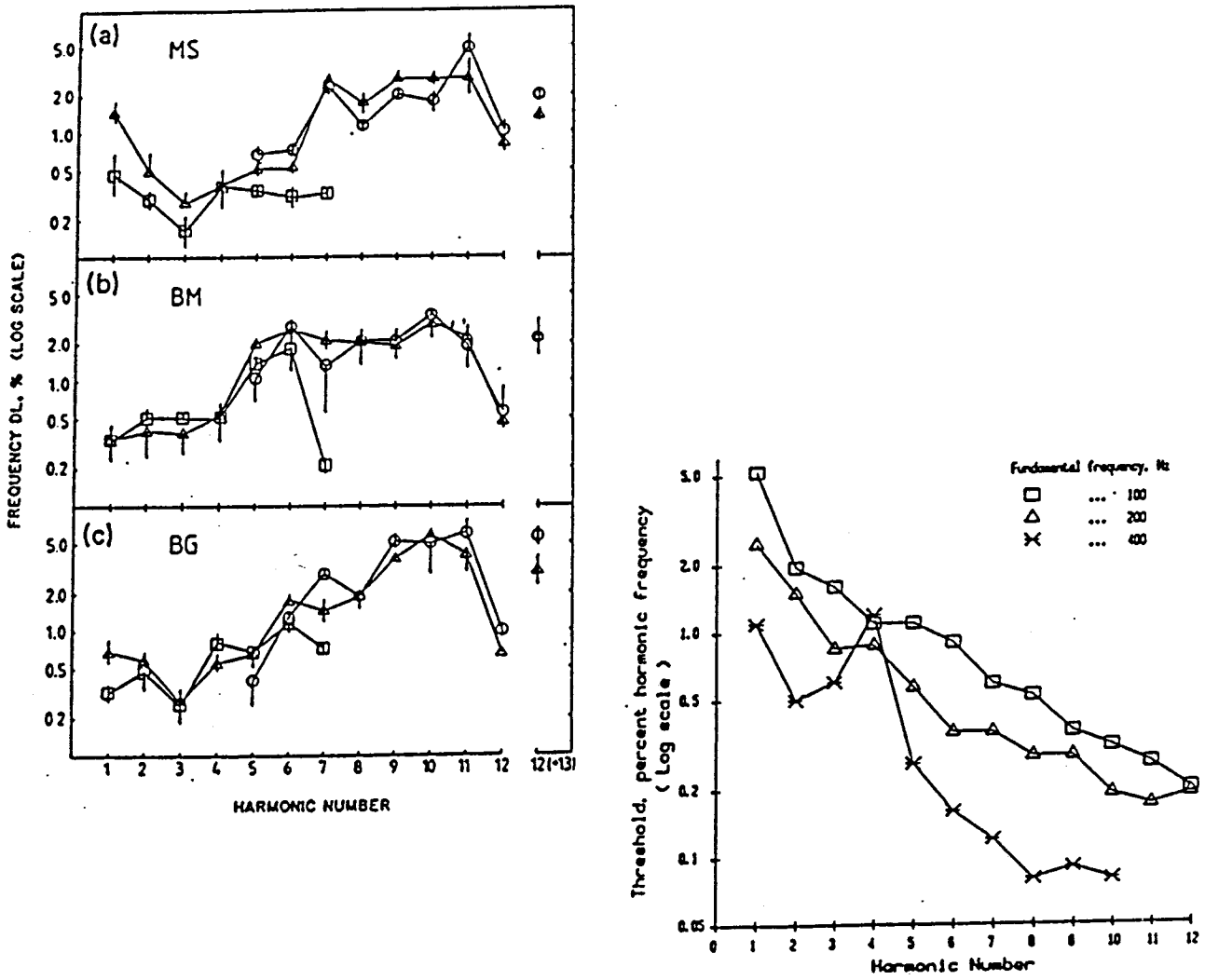


Figure 2.10 Results from two experiments by Moore et al. The top frame shows frequency DLs for individual components obtained with a *"pitch discrimination"* task, as a function of harmonic number. Three types of complex stimuli differed in the numbers of harmonics: 1-7 (squares), 1-12 (triangles), 5-12 (circles). The subjects were the authors. [From Moore et al. 1984]. The bottom frame shows average thresholds for *"detection of inharmonicity "* (same three subjects as above). Stimuli were complex tones comprising 10 or 12 harmonics. A single harmonic was "mistuned", as it was for the DL experiment represented above. [From Moore et al., 1985 b]

Changes in frequency of n=1-6 were found to influence the overall "residue" pitch to a greater degree than changes in higher harmonics. The shift in residue pitch was a linear function of the shift in the frequency of the mistuned harmonic for $\Delta f \leq \pm 2\%$ to3%. For greater "mistunings", the shift in residue pitch was *reduced,* rather than increased. Beyond the 2%-3% range, the mistuned component appeared to be ignored in pitch estimation.

This result led Moore et al. (1986) to wonder if this value of Δf was a good estimate of the "mesh size" of the hypothetical "harmonic sieve" for pitch proposed by many "template"-type models (Duifhuis et al., 1982; Scheffers, 1983; Grandori, 1984). A sieve with acceptance limits bounded by the $\pm 2\%$ to 3% shift may reject components with greater deviations as not belonging to the complex. This idea of harmonicity being a cue for fusion and deviation from harmonicity being a cue for segregation was discussed in the earlier section (2.4.1) on perceptual fusion.

Moore et al. (1986) "attempted to measure more directly the degree of mistuning required to make a partial audible as a separate tone". Again, a stimulus similar to that used in the former experiments was used; namely, a multicomponent complex tone (12 harmonics), with one component mistuned in frequency. A single sound was presented on a trial, and listeners were asked to report if they heard "a single sound with one pitch or two sounds - a complex tone and a component with a pure-tone quality not belonging to the complex".

Thresholds obtained for components to be heard out as individual entities are shown in **figure 2.11**, along with data from two of the earlier experiments shown for comparison.

In general, the threshold mistuning required to hear out partials was larger than the DL values obtained for the "pitch" discrmination experiment and the thresholds for detection of "inharmonicity". Also, the thresholds for hearing out components increased for higher harmonic numbers. The DLs also showed a rising trend with increase in harmonic number, whereas thresholds for judgments of inharmonicity decreased.

Again, the difference in results appears to be a direct consequence of the difference in task. The discrimination of higher components appears to be mediated by timbral cues such as changes in "roughness" for the inharmonicity-detection task. For the hearing-out task, listeners were explicitly asked to avoid using such cues. A greater magnitude of shift was apparently necessary for subjects to be able to decide if one or two sounds were present. The DL data could be confounded with any of these cues, since the task assigned was inappropriate and feedback was provided. Listeners may have used cues in addition to pitch change in discriminating the test sounds.

Based on the cumulative results of the four experiments reviewed above, and the observation that even a component deviant enough in frequency to be audible as a separate tone (1.3%-2%) could contribute to the pitch of the complex (for mistunings up to 3%-4%), Moore et al. (1986) suggest that the "harmonic sieve" proposed in many models for

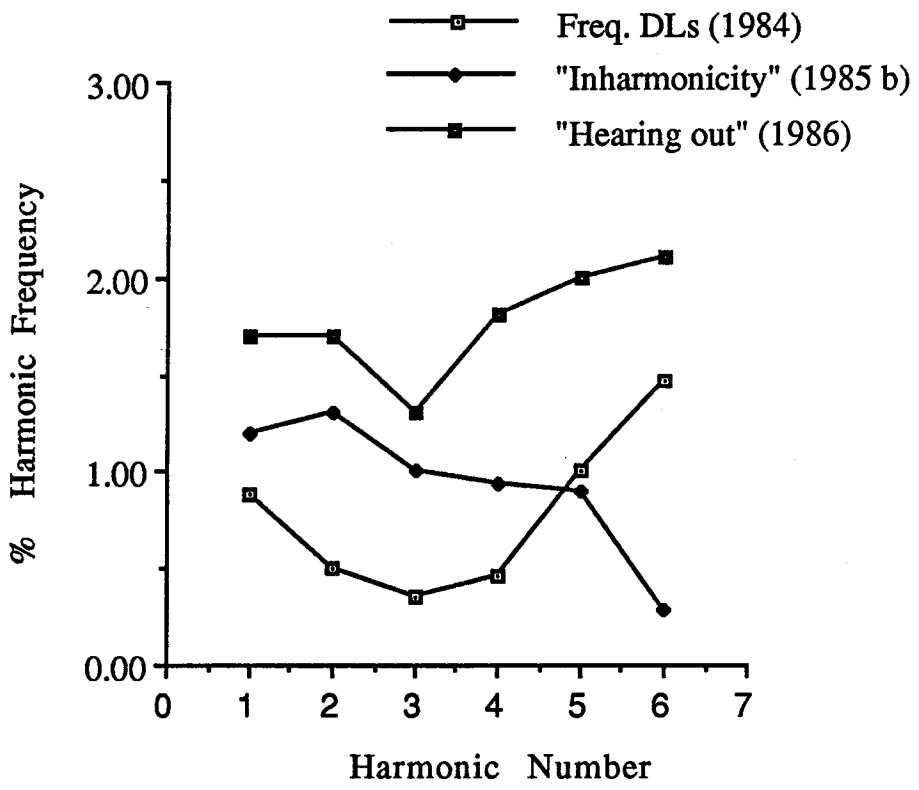


Figure 2.11 Thresholds for "hearing out" a mistuned component, for "detection of inharmonicity", and the component frequency DLs shown for the first 6 harmonics of 200 Hz. The same basic stimulus evoked different responses depending on the task assigned. [Adapted from Moore et al., 1984, 1985 b, 1986]

pitch may not have an "all-or-none" character. "Rather, a component may be weighted less and less in the pitch-determining process as it becomes more and more mistuned. At the same time the component would become more and more audible as a separate tone" (p. 482).

The **multiplicity of percepts** related to mistuning of harmonics in the experiments of Moore et al. highlights the point that **identifiable cues** can be used by listeners in making what on the surface appear to be generically opaque discrimination judgments. Differences in task and instructions can lead to differences in the selection of cues used.

2.11 Restatement of aim of dissertation

The many aspects of frequency discrimination reviewed above indicate that changes in the frequency of one or more components of complex tones can evoke **a range of percepts**, depending on factors such as the magnitude and location of the changes, context of stimulus presentation, and the nature of the task assigned to the listener.

The lower frequency DLs obtained for F0 of complex tones than for pure tones of equivalent F0 have been attributed to spectral pitch matches between frequencies of "corresponding" harmonics (Faulkner, 1985). However, discrimination of changes in the F0 of harmonic complexes appears to be affected by timbre differences related to the absolute frequency of components as well (Risset, 1971; Ritsma, 1963; Smoorenburg, 1970; Singh, 1987). The issue of pitch comparison being

affected by timbre difference and the role of corresponding harmonics in facilitating discrimination was the focus of experiment 1, reported in the next chapter.

Changes in the frequency of **all** components of a harmonic complex by the same linear amount are known to evoke changes in "residue" pitch (de Boer, 1956/1976; Schouten et al., 1962). McAdams (1984 b) reported that such stimuli were often perceived as comprising "multiple sources", but accompanying changes in pitch and timbre changes were not studied systematically. Experiment 2 took on the task of obtaining multiple perceptual impressions of such stimuli, both with regard to their perceived fusion, as well as with regard to their pitch and timbre.

For harmonic complexes in which a **single** harmonic is mistuned, divergent data on "thresholds" for discrimination exist. Hartmann (1989) claims that the threshold function is fairly flat across the range of harmonics, for hearing out of the mistuned component. Moore et al. (1984, 1985 a, b, 1986) obtained thresholds of different magnitude and differing functional dependence on harmonic number, depending on the task assigned. Experiment 3 was designed to integrate several choices into a single task, thereby providing a range of response "labels" to listeners, with which to report changes in perceived fusion and in the pitch and timbre of the changed sound.

Changes made in the frequency of one or more components of a complex sound possess the potential to yield **perceptual changes**

encompassing differences in perceived **pitch, timbre and fusion** of the altered complex. The experiments comprising this dissertation aimed to determine the perceptual correlates guiding discrimination judgements of spectral differences. The task and stimuli were designed specifically to enable exploration of the range of percepts and inference of their relation to physical stimulus features such as the **magnitude** and **location** of frequency changes and the **context** in which the changes were made.

The larger issue unifying the three experiments pertains to principles of grouping and segregation operative in facilitating fusion or fission, and the relation of such organizational processes to the perception of pitch and timbre. The results of the experiments are viewed in keeping with this broader perspective.

CHAPTER 3

EXPERIMENT 1

The influence of competing pitch and timbre cues on discrimination of missing fundamental frequency

3.1 Introduction

As discussed in the last chapter (section 2.9), there are several situations in which listeners have difficulties separating timbre from pitch. Nowhere is this conflict as explicitly manifested as in the case of "residue" tones (Schouten, 1940). For such tones, the overall residue pitch is usually considered to be synonymous with the fundamental frequency of the complex, even though a physical component at the fundamental frequency may not be present in the spectrum. For harmonic stimuli of this type, the periodicity of the waveform coincides with the frequency of the fundamental. Thus, a "period" cue is available, even though no "place" cues may be present to indicate a pitch corresponding to fundamental frequency.

The separability of place and period for residue tones has rendered them invaluable to experiments in which the *relative* contributions of place and period are the focus of investigation (Plomp, 1967; Ritsma, 1967) or where the ability to match spectral pitches of components is compared with the ability to match the overall, virtual pitch of the complex (Faulkner, 1985; Smoorenburg, 1970).

In order to avoid confounding comparisons of component pitches in studies that aimed to estimate the accuracy of *residue* pitch discrimination, some researchers attempted to obtain pitch matches or pitch-difference judgments between test sounds and reference sounds that had no **"corresponding"** or **"coincident"** components (i.e. their respective harmonics had different numerical ratios to the fundamental).

Thus, for example, Schouten et al. (1962) used a pair of complex sounds comprising harmonics (n= 6, 7, 8) and (n= 9, 10, 11) respectively. Ritsma (1963) used test and reference signals with harmonics (4, 5, 6) and (7, 8, 9) (among others). Faulkner used signals comprising harmonics (1, 2, 7, 8, 9) and (4, 5) or (5, 6). The lack of corresponding harmonics, was hoped to force listeners to attend to the "true" residue pitch, rather than to spectral pitches asscociated with individual components (Faulkner, 1985).

The results of pitch discrimination experiments with stimuli of this type invariably led to higher DL values for F0, than were found for pure tones or complex tones of equivalent spectral composition. Ritsma (1963) and Faulkner (1985) alleged that the lower DLs obtained for complex tones with "corresponding" components were indicative of discrimination of *components* of a complex, rather than of its' residue pitch. The integration of information across harmonics aided discrimination, thereby resulting in lower overall DLs.

In defense of several discrimination experiments (reviewed in sections 2.10.4, 2.10.5 and 2.10.6), Moore et al. (1984, 1987) discounted

this allegation, attributing the discrepancy in DLs obtained for tones with and without corresponding harmonics, to the "marked difference in timbre" in the latter case.

Timbre is strongly related to the position and shape of the spectral envelope. A change in timbre may thus ensue if spectral frequencies of components are changed. While experimenters using residue stimuli comprising different harmonic numbers acknowledged their different timbres, the possible influence of timbre on the results obtained was not considered. Pitch comparison was assumed to be independent of timbre difference.

The experiment reported here was motivated by the desire to determine what perceptual cues are used by a listener in making discrimination judgements for residue-type tones, and to specifically address the issue of possible **confounding of pitch and timbre**. The stimuli and task have been especially designed to allow testing of the **twin hypotheses** that corresponding harmonics aid discrimination, and that timbre differences impair pitch comparison.

3.2 Stimuli :

A "stimulus" comprised two presentations of 2 harmonic complex tones presented sequentially. All tones were synthesized digitally by adding 4 components with equal amplitudes. The tones were 100 msec in duration with 10 msec rise/fall times and a raised-cosine envelope. A

300-msec silent interval separated the 2 sounds in the sequence. . The sequence was presented twice per trial, with a 1600 msec repetition interval at a sound pressure level of about 70 dB.

The two tones in a sequence were either identical, or made to differ by changing either the spectral spacing (i.e. F0), or the spectral locus of components, or both spacing and locus simultaneously. These manipulations allowed interplay between two aspects of frequency known to be correlated with pitch and timbre (Schouten et al., 1962; Singh, 1987). For harmonic complex tones, the inter-component spacing equals the F0 and may theoretically be considered to define "pitch". Similarly, timbre, which is related to the location of spectral energy (Risset and Wessel, 1982) may be "operationally" defined for these stimuli by the locus of components as shown in **figure 3.1.**

The term "timbre" for a spectral locus defined by Lm, and the term "timbre contrast" for a locus relation defined by (Lm-Ln) are sometimes used to imply the relation between spectral locus and timbre. However, it should be remembered that timbre is really an aspect of perceptual experience, while the parameter "spectral locus", refers to a physical aspect of the stimulus.

The experimental parameters varied across trials were the F0 of the second tone, and the spectral loci of the two tones. The first tone in a sequence served as a "standard" tone for comparisons of F0. The F0 of this first tone was fixed to be either 200 or 400 Hz, while changes were made in the F0 of the second tone relative to the reference F0.

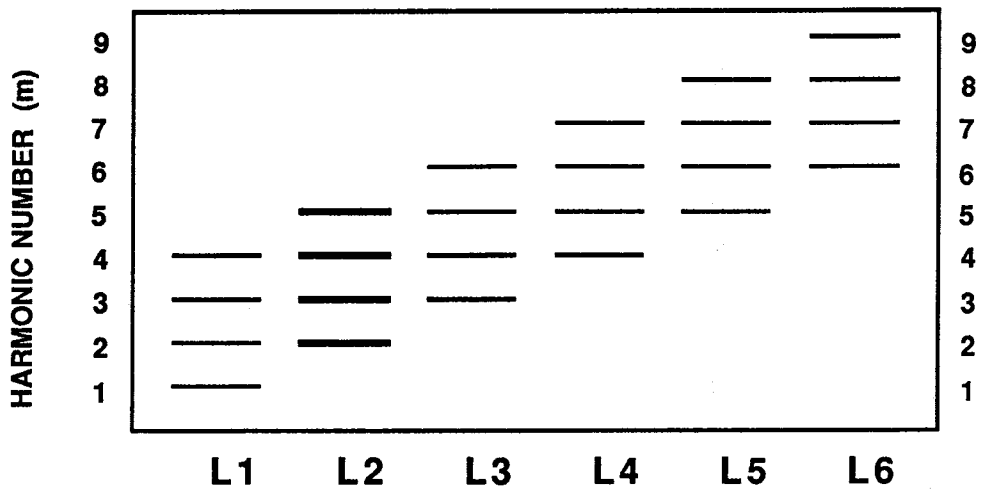


Figure 3.1 Spectral design for creating tones of different timbres: The locus *"Lm"* of four equal-amplitude harmonics *m, m+1, m+2* and *m+3* was varied to provide timbre variation (m=1,2,3,4,5, or 6). The six different loci were contrasted, pairwise, in sequences of the type (Lm-Lm), (L2-Lm) or (Lm-L2). Locus L2, often used as a reference, had three components in common with L1 and L3, two in common with L4, one with L5, and no components "corresponding" to those of L6.

The second tone could have the same F0 as the standard, or a higher or lower F0. The increment or decrement in F0 was selected from the range (200±2n) Hz or (400±4n) Hz, where n=0, 1, 2, 4, 8, 16, or 32, giving thirteen different values of F0 for the second tone at each reference frequency.

The spectral locus of tones in a pair was similarly made to be "same" or "different". The two tones were sometimes assigned the same spectral locus, giving an iso-locus sequence (Lm-Lm). In other cases, the locus of the second tone was changed to give a contrasted sequence of the type (Lm-Ln) or (Ln-Lm).

In sequences where locus changed across tones of the pair, **"L2"** was often used as a reference (i.e. n=2 in (Lm-Ln)). This particular locus was selected as a reference, because of its unique relationship with the other defined loci. The spectral locus L2 commencing with the second harmonic, has three components in common with the loci L1 and L3, two components in common with locus L4, one component in common with locus L5, no components in common with locus L6, and all components in common with itself. The sequence design thus provided a greater range in which to explore the importance of "corresponding" harmonics in discriminating F0 for residue tones (Faulkner, 1985).

Two examples of stimulus sequences constructed in this way are shown in **figure 3.2**. The top part of figure 3.2 illustrates a sequence in which the locus changes from L2 to L1, but F0 (and thus spectral spacing) remains the same (200 Hz). The bottom part illustrates a sequence in

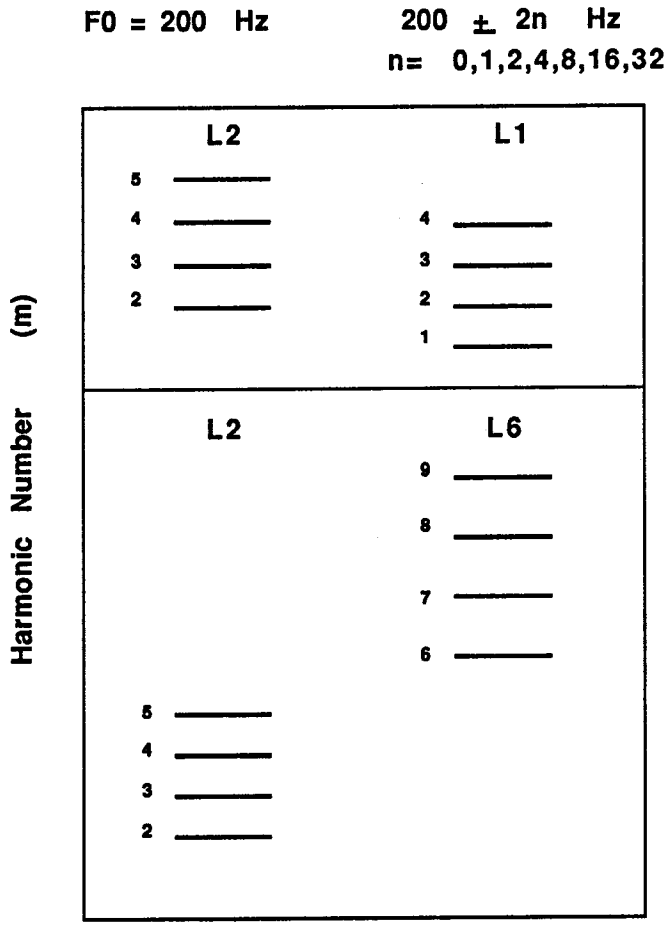


Figure 3.2. Two examples of sequences used in experiment 1: Thetop frame depicts a sequence in which the two tones had the same F0 (=200 Hz), but different spectral loci L2 and L1. The bottom frame depicts a sequence in which both the F0 and the spectral locus (L6) of the second tone are higher in frequency relative to the first tone (L2). Rise in F0 (from 200 to 200+Δf Hz) is manifested by the wider spacing of components of the second tone.

which there is an ascent in spectral locus from L2 to L6 as well as a rise in F0 (manifested by increased spectral spacing of components).

The locus L2 was used as a reference in both the first and second positions of contrasted sequences. Thus, for each type of locus condition ((Lm-Lm), (L2-Lm) or (Lm-L2)), there were 6 contrasts corresponding to m=1-6. Of these 18 contrasts (3 contrasts X 6 values of m), only 16 were unique, since the pair (L2-L2) was redundant across the 3 conditions (Lm-Lm), (L2-Lm) and (Lm-L2).

For each of the 16 contrasts of spectral loci, there were 13 different sequences corresponding to the 13 possible F0 relations between the 2 tones, ranging from 0% change to $\pm$ 1,2,4,8,16,32 % change in F0. There were thus 16X13 = 208 sequences for one reference F0. Since two values of reference F0 (=200, 400 Hz) were used, the total number of sequences synthesized for this experiment was 2X208 = 416.

3.3 Apparatus :

All stimuli were generated digitally by additive synthesis on a PDP 11/73 clone computer operating at a sampling frequency of 20 kHz. A program entitled PSYACX (Lai et al., 1987) allowed components to be specified in terms of their frequency, phase, amplitude, duration and envelope characteristics. After digital-to-analog conversion, the stimuli were further band-pass filtered between 100 Hz and10 kHz via a

brickwall filter (Wavetek-Rockland 751A) with response rolling off by 115 dB/octave outside the passband. The stimuli were then amplified to a comfortable listening level (about 70 dB SPL) using digital attenuators in conjunction with power amplifiers, and were presented binaurally over headphones (AKG 141) to the ears of a listener seated in a sound-absorptive listening room. The accuracy of the synthesized spectra was verified via a Hanning-window analysis on a 2033 single-channel spectrum analyzer by Bruel & Kjaer. A Tektronix dual-trace oscilloscope was used to verify the accuracy of the temporal layout of the stimuli in a sequence.

3.4 Procedure :

3.4.1 Subjects

Six listeners, between the age of 21 and 41 years, were recruited as subjects via a job advertisement (3 male: GC, LC, JO, and 3 female: CS, LJ, WK). Their hearing was verified to be within normal limits. Two of the subjects (GC and WK) are professional musicians with advanced musical training. Subject JO is active in amateur music performance and subject LJ has some training in dance. Subjects CS and LC are avid listeners of music but did not receive formal musical instruction.

3.4.2 Task

Stimuli were presented to listeners in a randomized order selected from pre-determined subsets of the total inventory of sequences with different locus and F0 relations as defined above. The listener's task was to indicate whether the second tone in a pair was: (1) same, (2) higher in pitch, (3) lower in pitch, (4) same in pitch but different in "something else", (5) higher in pitch and different in "something else", or (6) lower in pitch and different in "something else" than the first tone.

The term **"something else"** was used *intentionally,* to cover those timbral percepts that elude ease of description, and is considered to be synonymous with a change in "timbre" (which was described to listeners as being a feature of sounds distinct from their pitch, loudness or duration). A visual cue sheet (illustrated in **figure 3.3**) was provided to aid listeners in remembering the response labels. The listener indicated the choice of label by pressing the number keys '1', '2', '3', '4', '5' or '6' on a computer keyboard in accordance with the 6 options available.

3.4.3 Rationale for design of task and stimuli

The six options allowed in the labelling task were selected to facilitate reporting of changes in both timbre and pitch, independently or simultaneously. The selection of labels could then be compared to the

stimulus design to identify features of the spectrum that were being attended to. The *a priori* expectation was that a change in F0 should evoke a change in pitch and a change in locus should evoke a change in timbre. "Theoretically correct" responses would thus be to select labels '2' or '3' to report changes in pitch when F0 changed, and to select labels '5' or '6' when both locus and F0 changed. If only locus changed, but not F0, label '4' would be the "correct" response. If neither locus, nor F0 changed, the tones were identical and '1' would be the correct response.

However, the **real situation** is not so simple or straightforward. The "theoretically correct" responses may not correspond to the perceptual experience of listeners. There are many indications that pitch and timbre may not be orthogonal to each other (see section 2.9). Changes in F0 could affect timbre and changes in locus could affect pitch as well. The basic purpose of the present experiment was to seek out such interactions. The options provided in the task enabled combined changes in pitch and timbre to be reported explicitly.

The choice of harmonics for different "spectral loci" and their juxtaposition in a sequence further allowed examination of the claim of Faulkner (1985), that only tones lacking corresponding harmonics are compared for residue pitch. Thus, a sequence comprising tones with loci L2 and L6 (which lack common components) with the same F0, should elicit use of label '4' to imply equality of residue pitch. If label '4' was not selected, but rather, labels '5' or '6' were chosen, the timbre difference would be indicted as having influenced the judgment of pitch.

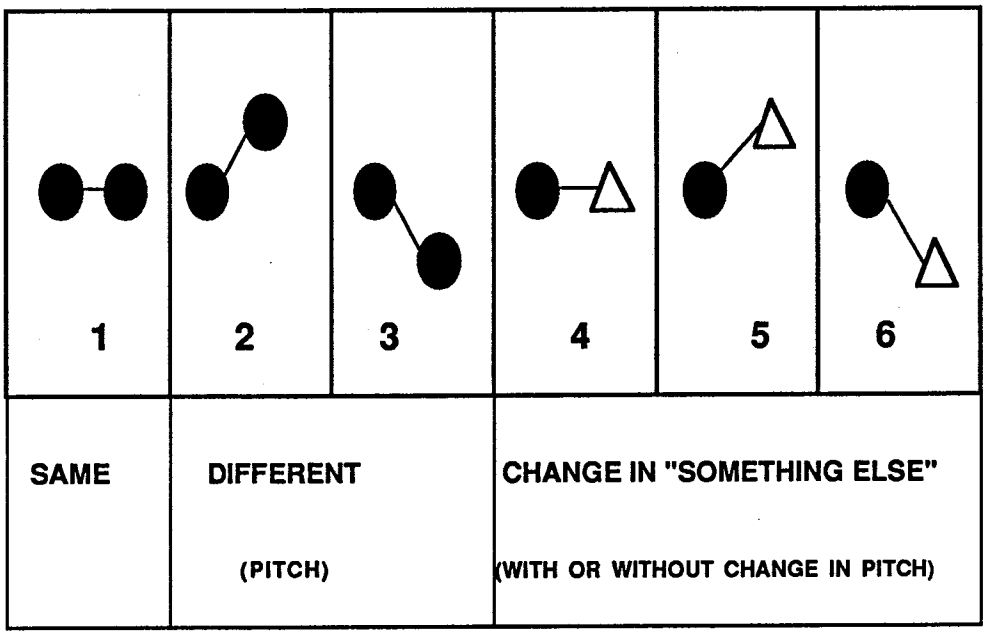


Figure 3.3 Illustration of the 6-option labelling task designed to to enable reports of pitch and timbre differences between the two tones of a stimulus sequence. Listeners had to indicate whether the second tone in a pair was: (1) same, (2) higher in pitch, (3) lower in pitch, (4) same in pitch but different in "something else", (5) higher in pitch and different in "something else", or (6) lower in pitch and different in "something else" than the first tone.

The claim of Moore et al. (1984; 1987) that timbre difference confounded judgments of pitch in the experiments of Ritsma (1963) and Faulkner (1985), would be verified, while Faulkner's claim that residue pitch is compared for such stimuli *uninfluenced* by component comparisons, would be rejected.

For loci with overlapping components ((L2-Lm) or (Lm-L2), with m=1, 2, 3, 4, 5), comparison of F0 and locus frequencies made in the same direction would provide convergent information for pitch change. Labels '5' or '6' thus ought to be used correctly. If, however, the F0 and locus frequencies changed in *opposite* directions, and listeners chose the label conveying the direction of F0 change, they could be construed as attending to residue pitch. If, instead, they reported a pitch change in the direction of the locus change, then multiple component comparisons would be indicated. The locus change may further be correlated with a difference in perceived timbre.

3.4.4 Stimulus presentation

The stimuli were subdivided into two sets corresponding to the two standard fundamental frequencies (200 and 400 Hz). The 200-Hz condition was run first.

Twelve stimuli, counterbalanced for order of change in locus at F0=200 Hz and F0=(200$\pm$2n) Hz, were presented four times each in a randomized order in a 48-trial "block". A particular value of the lowest

harmonic "m" defining a locus Lm was used as one of the parameters differentiating different "blocks" of stimuli. Thus, a block included stimulus sequences having four timbre-locus relations (L2-L2), (Lm-Lm), (L2-Lm) and (Lm-L2) between tones, for three values of frequency change ($\Delta F0=0$ and $\pm 2n$ Hz) to give a total of 4X3=12 stimuli per block. Like the locus rank "m", the F0 shift factor "n" was also used as a parameter differentiating blocks. A total of 36 blocks (6 locus contrasts X 6 differences in F0) were thus presented.

The program PSYACX was also used to run the experiment. In addition to administering the stimuli, it also kept a trial-by-trial record of the particular stimuli presented and the response keys selected by subjects. At the end of a block, the data could be retrieved as a matrix showing the set of stimuli presented in the block and the proportion of times different response keys were selected across replications.

The procedure was repeated for the 400-Hz condition.

3.5 Results :

The number of times particular response labels were selected to describe the stimulus presented across all 4 replications constituted the raw data. Thus, if a listener chose label 2 on 3 trials out of the 4 replications of a sequence, the response label 2 received a score of 3/4 or 75%. For each listener, a (6X12) matrix of responses was available that gave the proportions of the 6 possible responses (expressed as %

trials) for each of the 12 stimuli presented in a block. These data have been processed further in three different ways to facilitate discussion of the major trends observed:

i) Summary overview of data

The response matrices were examined for each listener and the label used most frequently in response to a particular stimulus was determined. This "label-of-choice" was that response label (of the 6 possible), that was used maximally and thus had the highest percentage across the 6 columns of a response matrix (in effect, this was the "mode" for each listener). Such "labels-of-choice" were collected for all listeners, for all stimuli. The label that was selected by the greatest number of listeners is considered to be the "dominant" representation of the stimulus being labelled. (This type of analysis has also been referred to as "modal analysis" (Singh, 1988) and "plurality of judgment" (Hawks, 1990)).

Such "dominant" response labels were derived for all locus contrasts for all values of F0 change for both reference F0s. These have been displayed in **tables 3.1 through 3.6** for F0=200 Hz to give a quick, summary view of the response trend for the entire stimulus set. The number in a single cell of the table corresponds to the response labels '1', '2', '3' '4', '5' or '6' that describe perceived changes in timbre and pitch as illustrated in the "key" accompanying each table. In cases where the distribution of response labels was multi-modal, i.e. more than

one label was selected by equal numbers of listeners, all the labels used are shown.

ii) Proportion of use of labels

To provide a quantifiable measure of the distribution of responses for a particular stimulus configuration, the proportions of use of individual labels have been averaged across subjects, while maintaining differences across response categories. To this end, 6 of the (6X12) response matrices (one for each subject) were averaged to give one mean matrix of responses. Each cell of the matrix gave the average number of times (in %) that a particular label (1,2,3,4,5, or 6) was used in response to a particular sequence characterized by a particular locus contrast and F0 relation between the two tones. The pattern of response selection is shown in detail in **figures 3.4, 3.5** and **3.6** for reference F0=200 Hz. Each frame in these figures represents a locus relation for a particular value of m in the context (Lm-Lm), (Lm-L2) or (L2-Lm). The abscissae give the magnitude of change in F0 (Hz) while the ordinates give the proportions of all of the labels used in terms of percentage of total numer of trials over all replications of the same stimulus. This type of representation, allows the distribution of proportion scores to be seen across labels, and is also useful in showing how the response selection changes as stimulus parameters (such as ΔF0 and m) change across the stimulus set.

iii) Distribution of response labels across listeners

The first scheme of data presentation gave a qualitative indication of the response label representative of the perceptual correlate of a stimulus change. The second scheme provided a quantitative measure of salience of use of particular labels, but entailed averaging across subjects. A third scheme, provides a blend of the previous two schemes in showing variability of label choice across listeners, that also reflects the salience of use of a particular label. In this case, the "labels of choice" for each stimulus have been plotted in terms of the number of listeners that selected them.

For purposes of graphic representation, a display scheme similar to that used in the previous graphs (figures 3.4, 3.5, 3.6) has been used, with the difference that the ordinate now represents the **number of listeners** favoring a particular response label, rather than the proportion of use of that label. The new graphs are shown in **figures 3.7, 3.8** and **3.9** for F0=200 Hz. Each figure shows responses for a different locus contrast (Lm-Lm), (L2-Lm) and (Lm-L2) and each frame within a figure corresponds to a particular value of harmonic number "m". As before, the data symbols '1', '2', '3', '4', '5' and '6' refer to the response label selected. (These numbers *should not be confused* with the "number of listeners", which is the variable represented along the ordinate).

Thus, if the label '3' was the one selected most frequently in response to a particular stimulus by 4 listeners and the label 6 was selected by 2 listeners, then the data symbol '3' will be plotted with a

y-axis value equal to 4 and the label '6' with a y-axis value equal to 2. Since the number of listeners was six, the limit of the ordinate is the number 6. However, it may be observed that in some cases, the numbers of listeners for a particular stimulus configuration add up to a number greater than 6. This occurs whenever there was any equivocation across response labels. Thus, if a listener was ambivalent and selected labels '4' and '5' equally often in responding to a particular stimulus, there were 2 modes of response, both of which have been included in the display.

These three methods of data representation have also been used to display the results obtained for reference F0=400 Hz. Those data are shown in **tables 3.7 through 3.12** and **figures 3.10 through 3.15.** Each of these representations is discussed separately below, for the two values of F0.

3.5.1 Data summary: Reference F0=200 Hz

Description of Tables 3.1 through 3.6 :

Table 3.1 shows dominant-label choice for upward changes in ΔF0 for the iso-locus sequences (Lm-Lm). Again, the "dominant" label refers to that label (of the 6 possible options provided in the task) that was chosen by the greatest number of listeners in response to sequences with locus relations as shown across the columns and F0 relations as shown down the rows.

TABLE 3.1 Summary table showing the label selected by the largest number of listeners in response to stimuli with locus relations as shown across the columns and F0 relations as shown down the rows. The labels correspond to perceptual features of the sounds as illustrated by the "key" shown below the table.

△ F0 (RE: 200 HZ)	LOCUS CONTRAST (Lm-Lm)					
	(L1-L1)	(L2-L2)	(L3-L3)	(L4-L4)	(L5-L5)	(L6-L6)
0	1	1	1	1	1	1
2	1 2	2	1 2	1 2	1 2	2
4	2	2	2	2	2	2
8	2	2	2	2	2	2
16	2	2	2	2	2	2
32	2	2	2	2	2	2
64	2	2	2	2	2	2

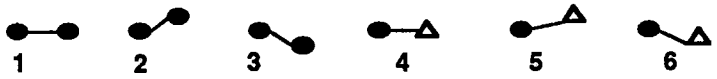
KEY: 

TABLE 3.2 Summary table showing the label selected by the largest number of listeners in response to stimuli with locus relations as shown across the columns and F0 relations as shown down the rows. The labels correspond to perceptual features of the sounds as illustrated by the "key" shown below the table.

△ F0	LOCUS CONTRAST (Lm-Lm)					
(RE: 200 HZ)	(L1-L1)	(L2-L2)	(L3-L3)	(L4-L4)	(L5-L5)	(L6-L6)
0	1	1	1	1	1	1
− 2	1 3	3	1 3	3	3	3
− 4	3	3	3	3	3	3
− 8	3	3	3	3	3	3
−16	3	3	3	3	3	3
−32	3	3	3	3	3	3
−64	3	6	6	3	3	3

KEY: 

TABLE 3.3 Summary table showing the label selected by the largest number of listeners in response to stimuli with locus relations as shown across the columns and F0 relations as shown down the rows. The labels correspond to perceptual features of the sounds as illustrated by the "key" shown below the table.

△ F0 (RE: 200 Hz)	LOCUS CONTRAST (L2-Lm)					
	(L2-L1)	(L2-L2)	(L2-L3)	(L2-L4)	(L2-L5)	(L2-L6)
0	6	1	5	5	5	4 5
2	6	2	5	5	4 5	5
4	6	2	5	5	5	5
8	5	2	5	5	5	5
16	5	2	5	5	5	5
32	5	2	5	5	5	5
64	5	2	5	5	5	5

KEY:

TABLE 3.4 Summary table showing the label selected by the largest number of listeners in response to stimuli with locus relations as shown across the columns and F0 relations as shown down the rows. The labels correspond to perceptual features of the sounds as illustrated by the "key" shown below the table.

△ F0 (RE: 200 Hz)	LOCUS CONTRAST (L2-Lm)					
	(L2-L1)	(L2-L2)	(L2-L3)	(L2-L4)	(L2-L5)	(L2-L6)
0	6	1	5	5	5	4 5
- 2	6	3	2 6	5	5 6	5
- 4	6	3	6	6	6	4 6
- 8	6	3	6	6	6	6
- 16	6	3	6	6	6	6
- 32	6	3	6	6	6	6
- 64	6	6	3	6	6	6

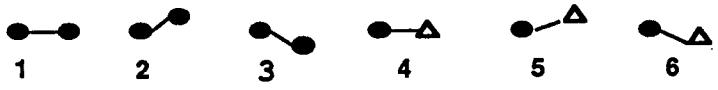
KEY: 

TABLE 3.5 Summary table showing the label selected by the largest number of listeners in response to stimuli with locus relations as shown across the columns and F0 relations as shown down the rows. The labels correspond to perceptual features of the sounds as illustrated by the "key" shown below the table.

△ F0	LOCUS CONTRAST (Lm-L2)					
(RE: 200 HZ)	(L1-L2)	(L2-L2)	(L3-L2)	(L4-L2)	(L5-L2)	(L6-L2)
0	5	1	6	6	6	6
2	5	2	4 6	6	6	6
4	5	2	6	6	6	6
8	5	2	5	6	5	4
16	5	2	5	5	5	5
32	5	2	2	5	5	5
64	5	2	2	2	5	5

KEY:

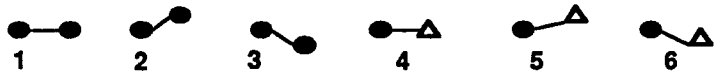


TABLE 3.6 Summary table showing the label selected by the largest number of listeners in response to stimuli with locus relations as shown across the columns and F0 relations as shown down the rows. The labels correspond to perceptual features of the sounds as illustrated by the "key" shown below the table.

△ F0	LOCUS CONTRAST (Lm-L2)					
(RE: 200 HZ)	(L1-L2)	(L2-L2)	(L3-L2)	(L4-L2)	(L5-L2)	(L6-L2)
0	5	1	6	6	6	6
− 2	5	3	6	6	6	6
− 4	6	3	6	6	6	6
− 8	6	3	6	6	6	6
− 16	6	3	6	6	6	6
− 32	6	3	6	6	6	6
− 64	6	6	6	6	6	6

KEY: 

The first row in table 3.1, for $\Delta F0 = 0$ Hz shows use of label 1 as would be expected for identical sequences ! The second row, with $\Delta F0 =$ 2 Hz shows use of labels 1 and 2. This value of $\Delta F0$ lies close to the ≈ 3 Hz jnd obtained for pure tones of equivalent F0 (Shower and Biddulph, 1931; Weir et al., 1977), so an ambivalence between same ('1') and different ('2') judgments is not surprising. For larger values of $\Delta F0$, listeners used the label '2' to indicate a rise in pitch without an accompanying timbre change. For these values of stimulus parameters, it seems that **spectral locus is indeed a good measure of timbre,** since tones defined to have the same locus were perceived as having the same timbre.

Similar results were obtained for iso-locus sequences with downward changes in F0, as shown in Table 3.2. Label '1' was used to indicate no change, while label '3' was used to indicate a downward change in pitch without an accompanying change in timbre. Two cells of this summary matrix are a little surprising; these correspond to $\Delta F0 =$ -64 Hz for locus contrasts (L2-L2) and (L3-L3). The dominant label chosen for these sequences was '6', rather than '3'. Listeners apparently heard "something else" change in addition to pitch for these sequences.

The stimulus design used here employed harmonics with equal amplitudes. This is an **unnatural** spectral design, given that most natural harmonic sounds exhibit a rolloff of amplitude in higher components (of about $1/f^3$ or around18 dB/octave beginning with the fifth or sixth harmonic (Benade, 1981)). Such a rolloff serves to reduce

inter-component masking in upper critical bands, "roughness" and other perceptually confusing effects, thus facilitating the tracking of pitch and tone color. The flat spectra used in this experiment may thus have been construed as "unnatural" or "tinny" for some values of spacing and locus.

It could also be that the reduced bandwidth of the flat spectral envelope with negative change in F0 for these particular spectral loci (L2 and L3) led to components interfering within critical bands. The frequency region spanned by these harmonics ranges from 400 Hz to 1200 Hz. Critical bandwidth for this range is between 100 and 200 Hz (Moore, 1982). A -64 Hz change leads to a spectral spacing (or F0) of 136 Hz. The higher components of the tones may therefore suffer such interference effects. Why this should occur only for the case (L2-L2) and (L3-L3) and not for other values of (Lm-Lm), however, is not clear at present.

Tables 3.3 and 3.4 show the dominant labels used in response to stimuli with locus relations (L2-Lm) as shown along the columns of the matrix and changes in F0 as shown down the rows. The data for the second column (L2-L2) are identical to those shown earlier in tables 3.1 and 3.2. The data in columns with locus relations defined by (L2-Lm, m>2) show use of label '5' to indicate a change in perceived pitch as well as a change in timbre. For sequences in which F0 did change (i.e. $\Delta F0 \geq$ 2Hz), this choice would be expected, since both locus and F0 change simultaneously.

However, the cross-hatched cells of the matrix showing use of label

'5' for sequences in which m > 2 and ΔF0=0 are surprising. The "expected" label for such sequences would be '4'. Since F0 did not change, the expectation based on pitch theories would be that the "residue" pitch (i.e. the overall pitch of the complex) should remain unchanged. However, most listeners appeared to perceive a rise in pitch, despite no change in F0.

The data shown with downward cross-hatching in the first column, for (L2-L1) are also divergent from expectations. For the sequences in these cells, listeners used label '6' indicating a **fall** in pitch, when F0 did not change (ΔF0=0), or in fact **increased** by ≤4 Hz. For greater increments in F0, the label choice changed to '5' indicating correctly perceived direction of F0 change.

The anomolous results described above seem to indicate that **decisions about pitch change are influenced not just by change in F0, but by change in spectral locus as well.** The ascent in spectral locus for sequences (L2-Lm, m>2) seems to have been construed as a rise in pitch, given the absence of other pitch cues. Similarly, the descent in locus for (L2-L1) is construed as a fall in pitch, even if F0 is increased, up until a limiting value of around 4 Hz.

Table 3.4 shows the dominant labels used for sequences with the same locus contrasts (L2-Lm) as table 3.3, but this time with a decrease in F0. The trend noted in table 3.2 is confirmed in this table. Listeners used label '5' for sequences with an ascent in locus (L2-Lm, m>2), and label '6' for a descent in locus (L2-L1), despite there being *no change* in

F0.

The use of label '2' for a -2 Hz change in the sequence (L2-L3) suggests that the downward change of 1% in the frequency of harmonics, with a simultaneous upward change to the adjacent harmonic may have led to a balancing of the opposing changes in a way as to minimize a locus-related difference in timbre. For even bigger downward changes in F0 ($\geq$ 4 Hz), the direction of F0 change was correctly perceived, as manifested by a perceived fall in pitch (use of label '6').

The use of label '4' for sequences with locus contrast (L2-L6) and a downward change in F0 of 4 Hz also points to a neutralizing effect. The perceived rise in pitch corresponding to the ascent in locus appears to be compensated by the downward change in F0 to give an overall result of no perceived pitch change, although a change in timbre was still perceived. Beyond this 4 Hz threshold, the direction of perceived pitch change corresponded to the direction of change in F0

The effect of direction of locus change on perception of pitch change was further verified for sequences with locus contrasts (Lm-L2). The dominant-label data are shown in tables 3.5 and 3.6. The influence of locus on perceived direction of pitch change is blatantly clear in the sequences (Lm-L2, m>2), shown cross-hatched in table 3.5. For F0 changes from 0 to +8 Hz, the use of label '6' rather than label '5', indicates that the direction of locus change dominated the perception of pitch. Thus, a descent in locus for these sequences, was construed as a fall in pitch, even though there was no change in F0, or, in fact, an upward

change !

For sequences (L3-L2), and (L6-L2) the use of label '4' is also seen for a +2 Hz change and a +8 Hz change, respectively. The reason for using label '4' could be similar to that proposed earlier for the use of label '2' in Table 3.4, namely that the perceived falling of pitch due to the descent in locus could be "neutralized" by the upward shift in the frequency of all harmonics, leading to listeners reporting no change in pitch, although a timbre change may still be perceived (use of label '4').

A similar type of compensation may be taking place for the +32 and +64 Hz change in sequence (L3-L2) and the +64 Hz change in sequence (L4-L2). These sequences were labelled '2', instead of the expected '5'. While the reported direction of pitch change (up) is congruent with the increase in F0, the absence of a timbre change seems to indicate that the 16 to 32% increase in frequencies of harmonics counteracted the descent in locus. This conflicting upward and downward pull in frequency may have left the tones of the sequence in basically the same spectral region. They were thus apparently perceived as having a similar timbre.

Table 3.6 shows dominant labels for decrements in F0 for the sequences (Lm-L2). The influence of direction of locus change is again seen as shown in the cross-hatched areas. However, in comparison with table 3.5, this effect is not as extensive across the range of F0 change. For increments in F0 (table 3.5), an influence of locus was seen for values as high as $\Delta F0 = 8$ Hz. In table 3.6, the effect is seen only for changes in F0

=0, and -2 Hz.

3.5.2 "Magnitude-of-use" of response labels (figures 3.4 through 3.7; Ref. F0=200 Hz)

The graphs plotted in figures 3.4 through 3.7 give a measure of the salience of a particular label as representative of the perceptual relations between tones of a stimulus sequence. The distribution of responses **across** the 6 posible options is also given in detail. Each frame shows the distribution of responses, plotted in terms of % trials, versus the change in F0 shown along the abscissa, for one locus contrast per frame. The data points are represented by the symbols '1', '2', '3',' 4', '5' and '6' in accord with the response label they represent. Since the response labels are plotted in terms of percentage of use, these graphs enable comparison of salience of response and also show the boundaries where one label is forsaken for another.

For the locus relation (Lm-Lm), **figure** 3.4 basically replicates the data already described in tables 3.1 and 3.2. These graphs show the dominant use of labels '1', '2' and '3' to indicate no change, or a pitch change without an accompanying change in timbre, as would be expected for harmonic complex tones having the same, general spectral locus.

The graphs for locus relation (L2-Lm) plotted in figure 3.5 validate the response trend described in the tables and also provide a quantitative measure of response in terms of percentage of use of the

different labels. For sequences in which there was an an ascent in locus (m>2), label '5' was selected on the greatest percentage of trials for a change in F0 ranging from +64 Hz down to -2 Hz. For sequences with a descent in locus (m=1), the label '6' was used to indicate a fall in pitch for F0 changes ranging from -64 Hz up to +4 Hz. It should be noted, however, that in all these cases, other labels were used as well, albeit not on as many trials as the dominant labels.

These graphs thus verify the trend noted from the tables that most listeners tended to follow the direction of locus change to indicate pitch change given a small difference in F0.

The main new fact observable in the graphs is the proportion of use of label '4'. For many of the sequences, this response label was the second most popular choice. However, its use was limited to $\leq 25\%$, with an increase as $\Delta F0$ approached 0 Hz. (As will be described later, this increase is due to the fact that some listeners did consistently choose this "correct" label for sequences in which $\Delta F0=0$. The high percentage contributed by their scores to these average data are manifested here in the level of use of label '4').

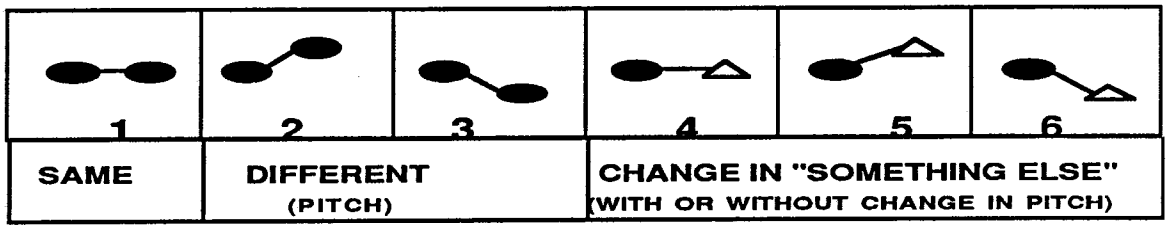
For decrements in F0 for the sequence (L2-Lm, m=4), the graph shows a 20 to 40% use of label 5. This again points to the fact that some listeners continued to follow direction of locus change (up), rather than direction of F0 change, in making judgments about pitch change.

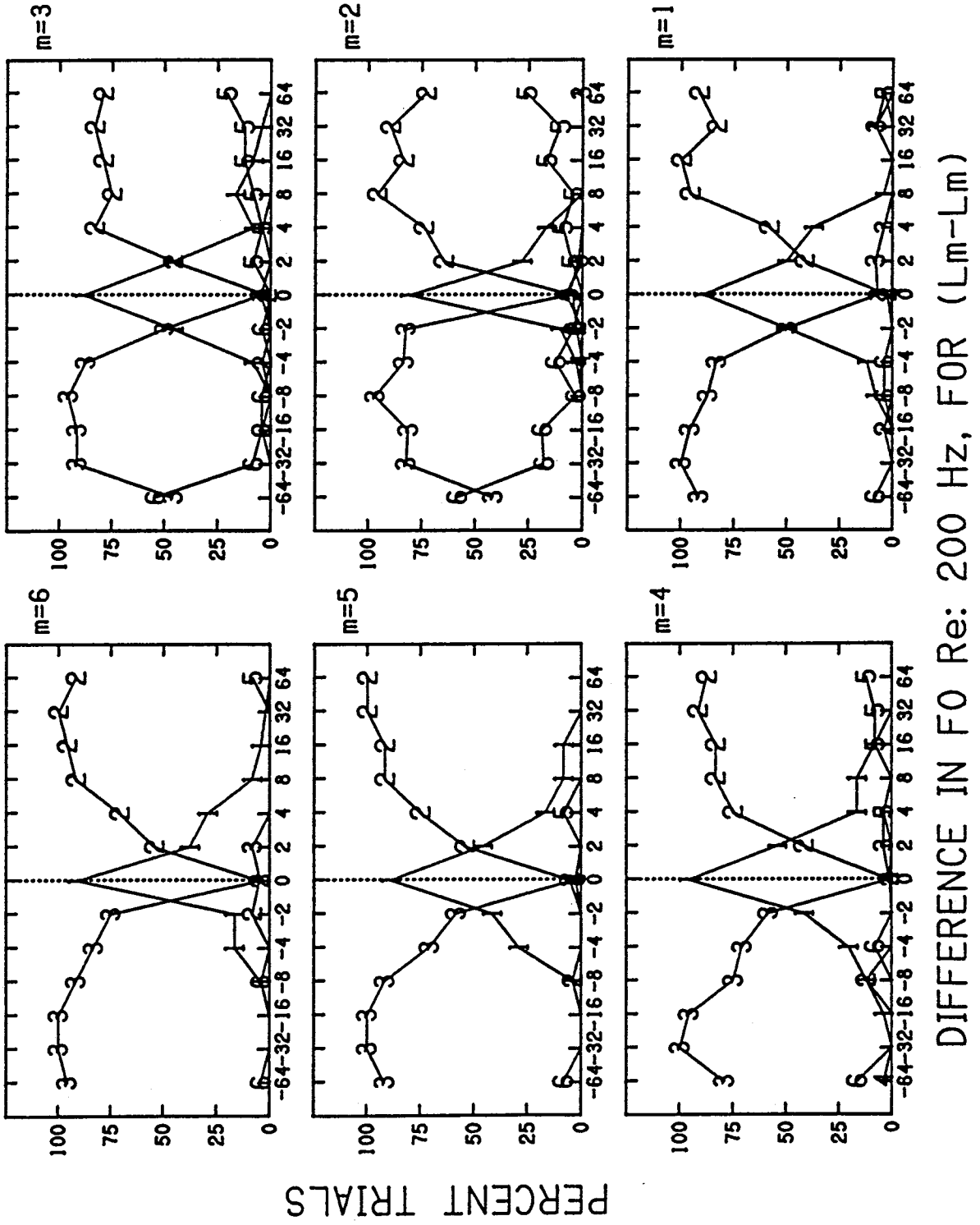
Figures 3.4 through 3.6 (next three pages)

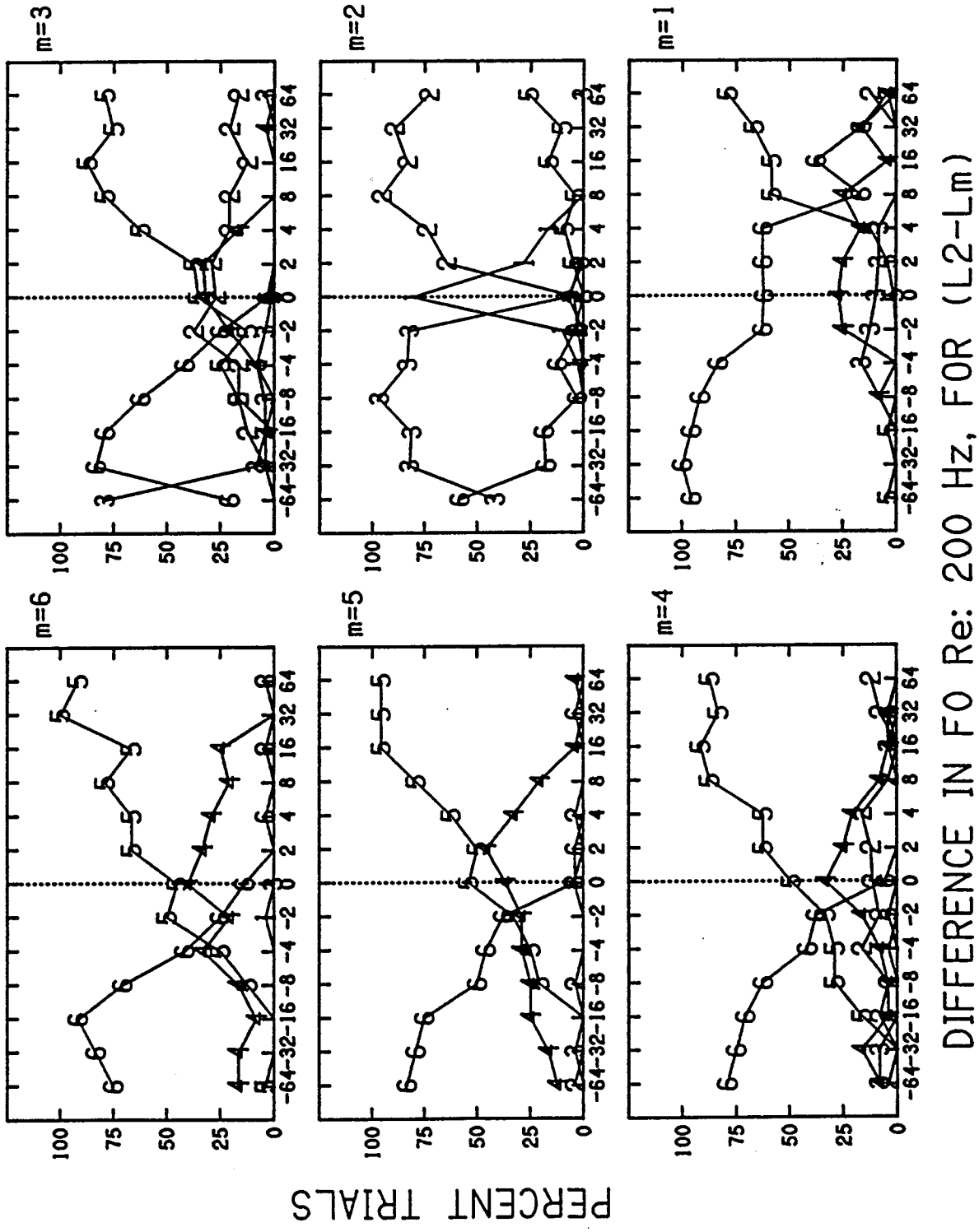
Each figure shows "magnitude-of-use" of different response labels averaged over 6 listeners, for locus relations between tones as mentioned at the bottom of the figure ((Lm-Lm), (L2-Lm) and (Lm-Lm), for figures 3.4, 3.5 and 3.6, respectively).

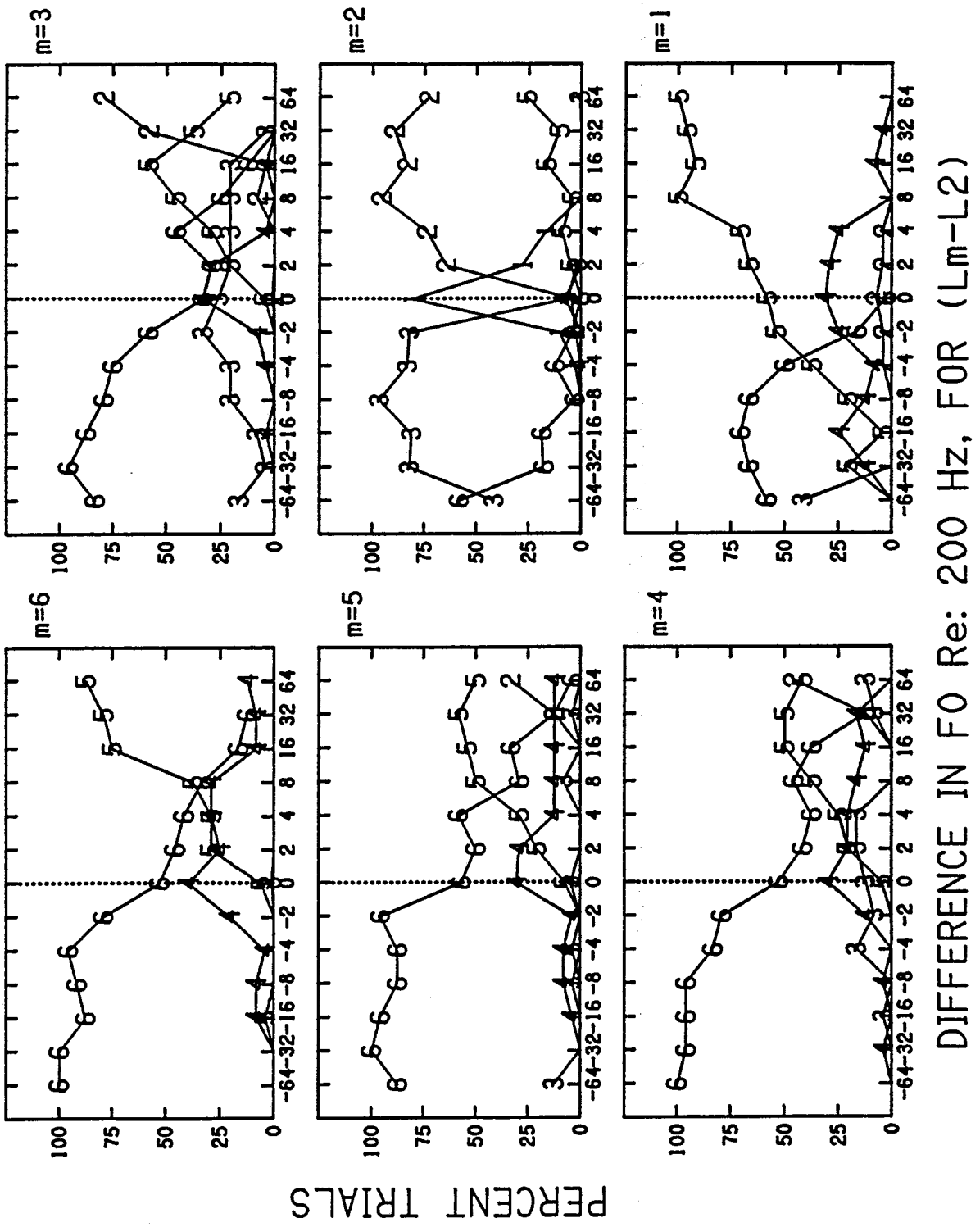
Each of the six "frames" corresponds to the value of the lower harmonic number "m" defining a locus. Differences in F0 between the two tones (re: 200 Hz) are displayed on the abscissa. The ordinate gives the proportion of trials on which different labels were used.

Numbers placed at data points correspond to the label number as defined in the task (illustrated below). Thus, '1' implies no change, '2', a rise in pitch without a change in timbre, '3' , a fall in pitch without a change in timbre, '4', a change in timbre without a change in pitch, '5' a change in timbre and a rise in pitch, and '6' a change in timbre and a fall in pitch.




m=3
m=2
m=1
m=6
m=5
m=4
PERCENT TRIALS
DIFFERENCE IN FO Re: 200 Hz, FOR (Lm-Lm)


m=3
m=2
m=1
m=6
m=5
m=4
PERCENT TRIALS
DIFFERENCE IN FO Re: 200 Hz, FOR (L2-Lm)


m=3
m=2
m=1
m=6
m=5
m=4
PERCENT TRIALS
DIFFERENCE IN FO Re: 200 Hz, FOR (Lm-L2)

The graphs for sequences with locus contrast (Lm-L2) plotted in figure 3.6 also show the use of dominant labels as given in tables 3.5 and 3.6, but with magnitude of response represented in terms of the proportions shown on the ordinate. One can thus follow the decreased usage of one label (e.g. '6'), as the proportions of other labels increase. The high percentage of use of label '6' for decrements in F0 seems to indicate that the perception of a fall in pitch was greatly enhanced by the mutually compatible cues of descent in locus accompanied by decrease in F0. This is further verified by looking at the graph for m=1, which shows less salient use of '6' for decrements in F0 and a more salient use of label '5', than for the other graphs. In this sequence, there is an ascent, rather than a descent in locus, which explains the reversing of the trend seen in the other graphs.

3.5.3 Listener-variability
(figures 3.7 through 3.9; Ref. F0=200 Hz)

The graphs in figures 3.7 through 3.9 correspond to the third type of data processing scheme described earlier. The number of listeners is represented along the ordinate, while the response label they selected is plotted as the data point. The abscissa gives F0 relations between the two tones of the sequence, while the locus relation is shown in terms of the "m" value for each frame.

At first glance, these three figures seem almost identical to the previously-described figures that showed proportion of use of different labels. This is a reassuring coincidence, indicating that listeners were in good agreement with each other in most cases. Thus, a high percentage of use of a particular label in the previous set of figures correlates positively with the high number of listeners selecting that label shown in the new figures.

For a 0 Hz change in F0, it can be seen that label '4' was used consistently by at least one listener. This particular listener (LC) was a good judge of periodicity change and reported a change in pitch only when there was a (supra-threshold) change in F0, regardless of direction of locus change. Another listener (WK) was at the other extreme. She consistently followed the spectrum, reporting a pitch change in the direction of locus change for small opposing changes in F0.

The other 4 listeners were less consistent. CS and LJ seemed more influenced by spectral change than F0 change, while GC and JO switched back and forth between following F0 change and reporting locus change as a pitch change for small shifts in F0. Individual differences across listeners are described in more detail in a following section.

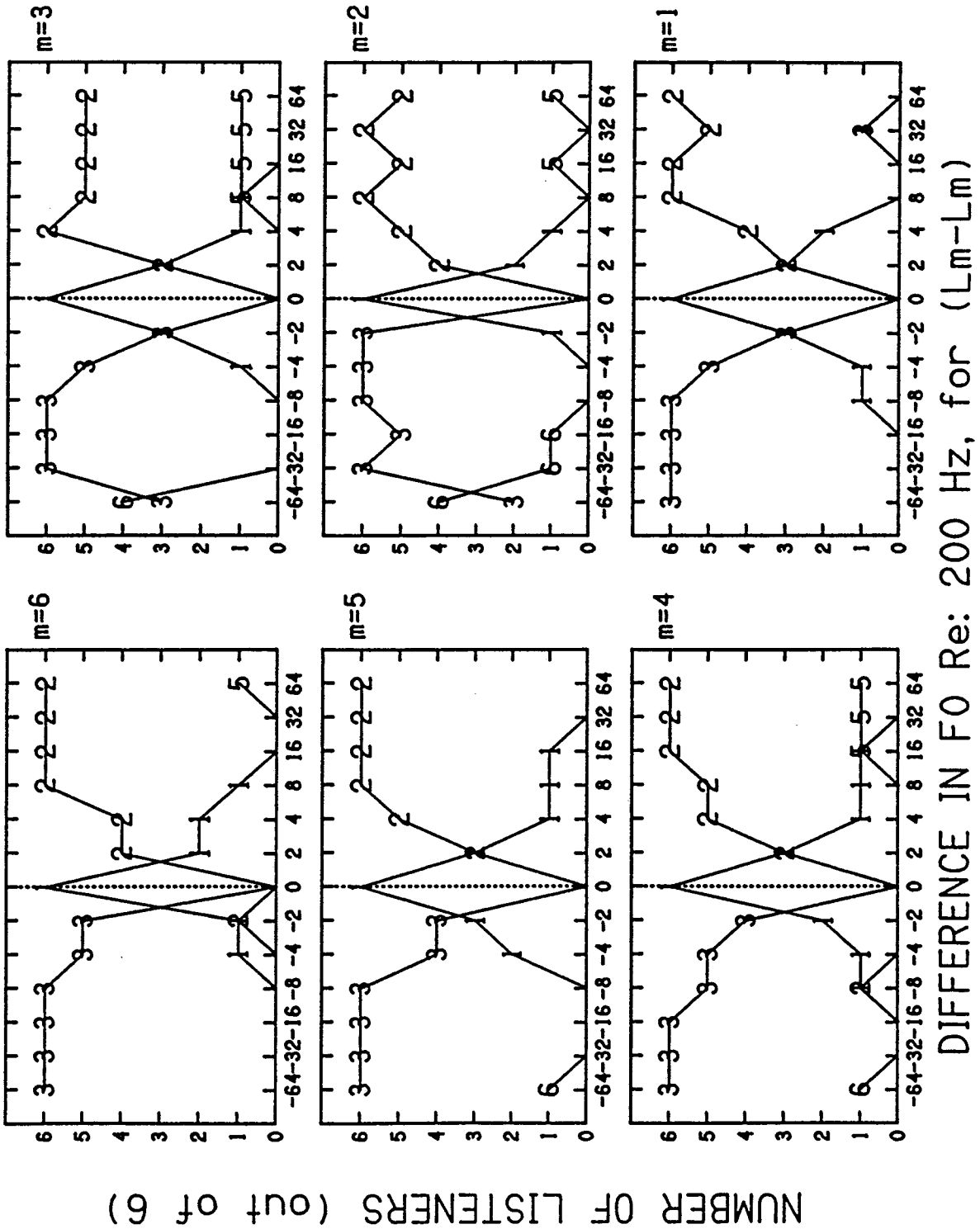
Figures 3.7 through 3.9 (next three pages)

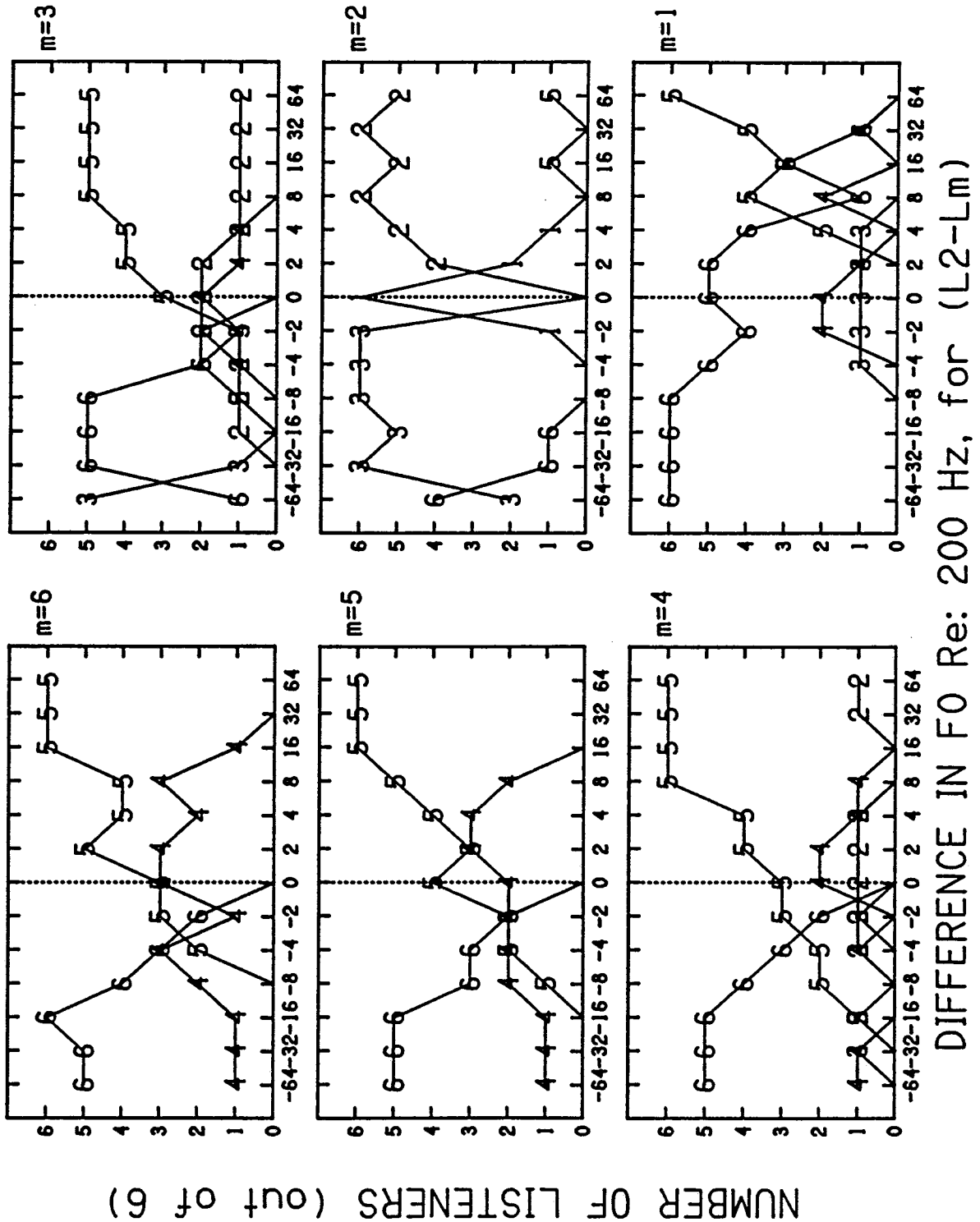
Each figure shows numbers of listeners and the "dominant label" selected by them for stimuli with locus relations between tones as mentioned at the bottom of the figure ((Lm-Lm), (L2-Lm) and (Lm-L2), for figures 3.7, 3.8 and 3.9, respectively).

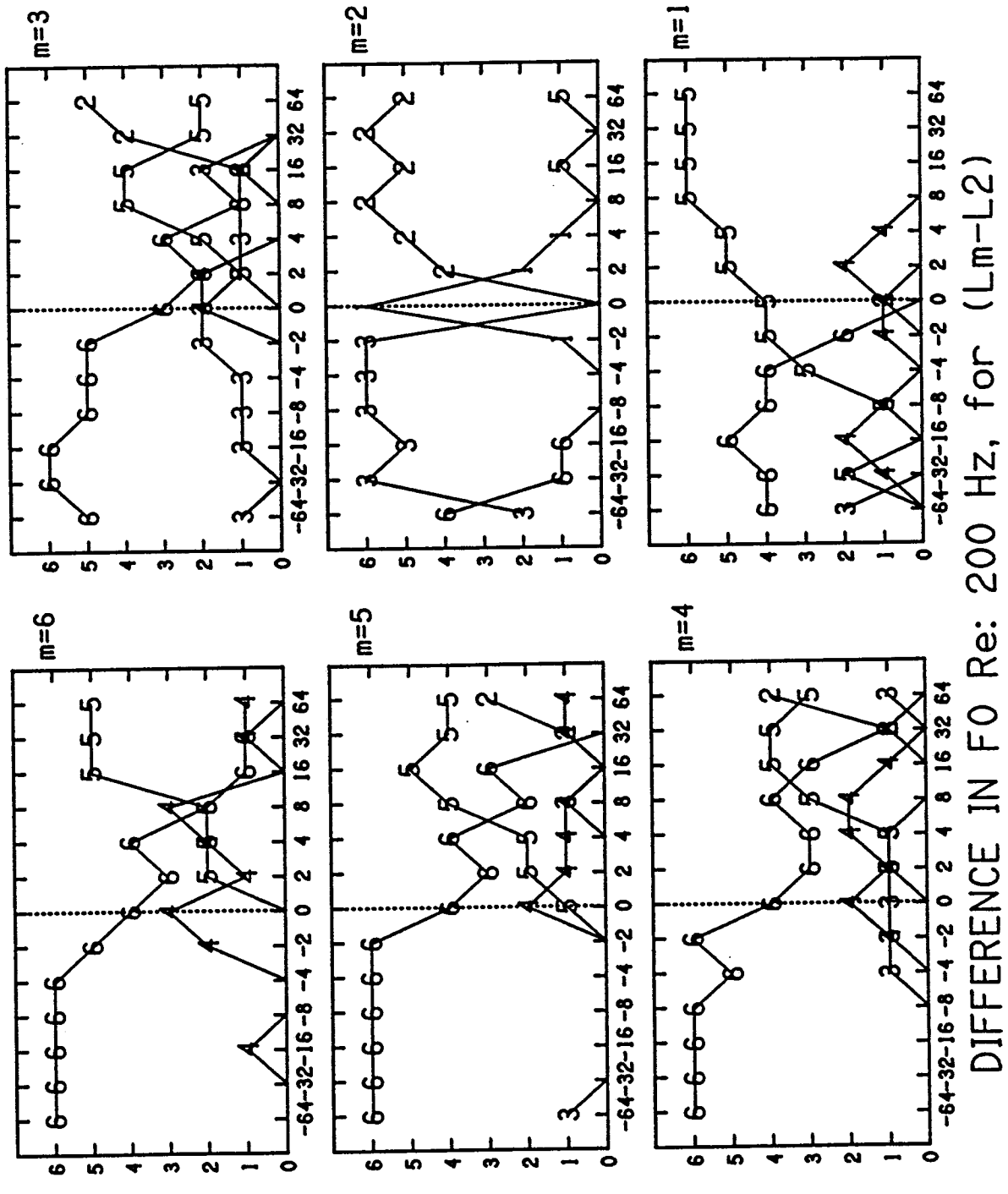
Each of the six "frames" corresponds to the value of the lower harmonic number "m" defining a locus. Differences in F0 between the two tones (re: 200 Hz) are displayed on the abscissa. The ordinate gives the number of listeners who selected the different labels that are plotted as data points.

As before, the data points correspond to the label number as defined in the task. Thus, '1' implies no change, '2' implies a rise in pitch without a change in timbre, '3' , a fall in pitch without a change in timbre, '4', a change in timbre without a change in pitch, '5' a change in timbre and a rise in pitch, and '6' a change in timbre and a fall in pitch.

Caution: Do not confuse the number of the response labels '1', '2', '3', '4', '5', '6' with the number of listeners (out of 6) represented along the ordinate.


m=3
m=2
m=1
m=6
m=5
m=4
-64 -32 -16 -8 -4 -2 0 2 4 8 16 32 64
0 1 2 3 4 5 6
DIFFERENCE IN F0 Re: 200 Hz, for (Lm-Lm)
NUMBER OF LISTENERS (out of 6)


m=3
m=2
m=1
m=6
m=5
m=4
-64 -32 -16 -8 -4 -2 0 2 4 8 16 32 64
0 1 2 3 4 5 6
DIFFERENCE IN F0 Re: 200 Hz, for (L2-Lm)
NUMBER OF LISTENERS (out of 6)


m=3
m=2
m=1
m=6
m=5
m=4
-64 -32 -16 -8 -4 -2 0 2 4 8 16 32 64
0 1 2 3 4 5 6
DIFFERENCE IN F0 Re: 200 Hz, for (Lm-L2)
NUMBER OF LISTENERS (out of 6)

3.6 Results for F0=400 Hz

Tables 3.7 through 3.12 (next six pages), and figures 3.10 through 3.15 show the results obtained with reference F0 of the first tone = 400 Hz. The second tone could be made to vary in F0 by an amount equal to $\Delta F0 = 4n$ Hz where $n = \pm 0, 1, 2, 4, 8, 16$ or 32. The data have been processed in the same three ways described earlier, as was done for F0=200 Hz .

The tables showing the labels selected by the largest number of listeners reveal **similar confusions** of spectral locus and perceived pitch change. However, the increased use of label '4' to indicate a timbre change unaccompanied by a pitch change for $\Delta F0=0$ Hz, indicates that listeners were better able to follow the periodicity of these signals as compared to the 200 Hz signals. This is also manifested in figures 3.10 through 3.15 by the dominance of label '4' at the 0 Hz value at the mid-point of the abscissae. In comparison with the 200-Hz stimuli, the confusions between direction of F0 change and direction of locus change were fewer, but they still did occur in the range $\Delta F0=0$ - + 2%.

TABLE 3.7 Summary table showing the label selected by the largest number of listeners in response to stimuli with locus relations as shown across the columns and F0 relations as shown down the rows. The labels correspond to perceptual features of the sounds as illustrated by the "key" shown below the table.

△ F0	LOCUS CONTRAST (Lm-Lm)					
(RE: 400 HZ)	(L1-L1)	(L2-L2)	(L3-L3)	(L4-L4)	(L5-L5)	(L6-L6)
0	1	1	1	1	1	1
4	2	2	2	2	2	2
8	2	2	2	2	2	2
16	2	2	2	2	2	2
32	2	2	2	2	2	2
64	2	2	2	2	2	2
128	2	2	2	2	2	2

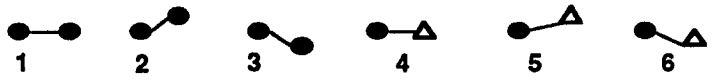
KEY: 

TABLE 3.8 Summary table showing the label selected by the largest number of listeners in response to stimuli with locus relations as shown across the columns and F0 relations as shown down the rows. The labels correspond to perceptual features of the sounds as illustrated by the "key" shown below the table.

△ F0	LOCUS CONTRAST (Lm-Lm)					
(RE: 400 HZ)	(L1-L1)	(L2-L2)	(L3-L3)	(L4-L4)	(L5-L5)	(L6-L6)
0	1	1	1	1	1	1
− 4	3	3	3	3	3	3
− 8	3	3	3	3	3	3
−16	3	3	3	3	3	3
−32	3	3	3	3	3	3
−64	3	3	3	3	3	3
−128	3	6	3	6	3	3

KEY: 

TABLE 3.9 Summary table showing the label selected by the largest number of listeners in response to stimuli with locus relations as shown across the columns and F0 relations as shown down the rows. The labels correspond to perceptual features of the sounds as illustrated by the "key" shown below the table.

△ F0 (RE: 400 Hz)	LOCUS CONTRAST (L2-Lm)					
	(L2-L1)	(L2-L2)	(L2-L3)	(L2-L4)	(L2-L5)	(L2-L6)
0	6	1	4	4	5	4 5
4	6	2	4 5	5	5	5
8	5	2	5	5	5	5
16	5	2	5	5	5	5
32	5	2	5	5	5	5
64	5	2	5	5	5	5
128	5	2	5	5	5	5

KEY:

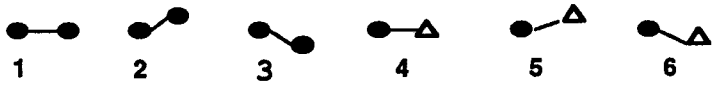


TABLE 3.10 Summary table showing the label selected by the largest number of listeners in response to stimuli with locus relations as shown across the columns and F0 relations as shown down the rows. The labels correspond to perceptual features of the sounds as illustrated by the "key" shown below the table.

△ F0 (RE: 400 HZ)	LOCUS CONTRAST (L2-Lm)					
	(L2-L1)	(L2-L2)	(L2-L3)	(L2-L4)	(L2-L5)	(L2-L6)
0	6	1	4	4	5	4 5
− 4	6	3	6	6	6	5
− 8	6	3	6	6	6	6
−16	6	3	6	6	6	6
−32	6	3	6	6	6	6
−64	6	3	6	6	6	6
−128	6	6	3	6	6	6

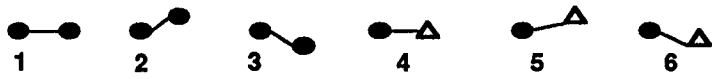
KEY: 

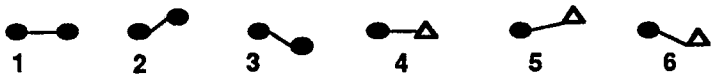
TABLE 3.11 Summary table showing the label selected by the largest number of listeners in response to stimuli with locus relations as shown across the columns and F0 relations as shown down the rows. The labels correspond to perceptual features of the sounds as illustrated by the "key" shown below the table.

△ F0 (RE: 400 HZ)	LOCUS CONTRAST (Lm-L2)					
	(L1-L2)	(L2-L2)	(L3-L2)	(L4-L2)	(L5-L2)	(L6-L2)
0	4 5	1	4	4	4	4
4	5	2	5	4	6	5
8	5	2	5	5	5	5
16	5	2	5	5	5	5
32	5	2	5	5	5	5
64	5	2	2	5	5	5
128	5	2	2	2	5	5

KEY:

TABLE 3.12 Summary table showing the label selected by the largest number of listeners in response to stimuli with locus relations as shown across the columns and F0 relations as shown down the rows. The labels correspond to perceptual features of the sounds as illustrated by the "key" shown below the table.

△ F0 (RE: 400 HZ)	LOCUS CONTRAST (Lm-L2)					
	(L1-L2)	(L2-L2)	(L3-L2)	(L4-L2)	(L5-L2)	(L6-L2)
0	4 5	1	4	4	4	4
– 4	6 5	3	6	6	6	4 6
– 8	6	3	6	6	6	6
– 16	6	3	6	6	6	6
– 32	6	3	6	6	6	6
– 64	6	3	6	6	6	6
– 128	6	6	6	6	6	6

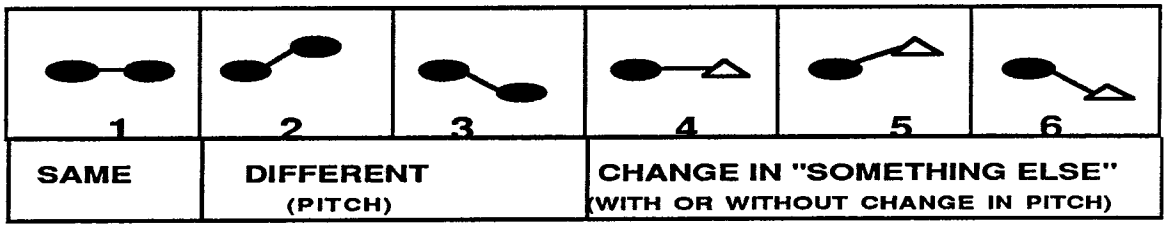
KEY: 

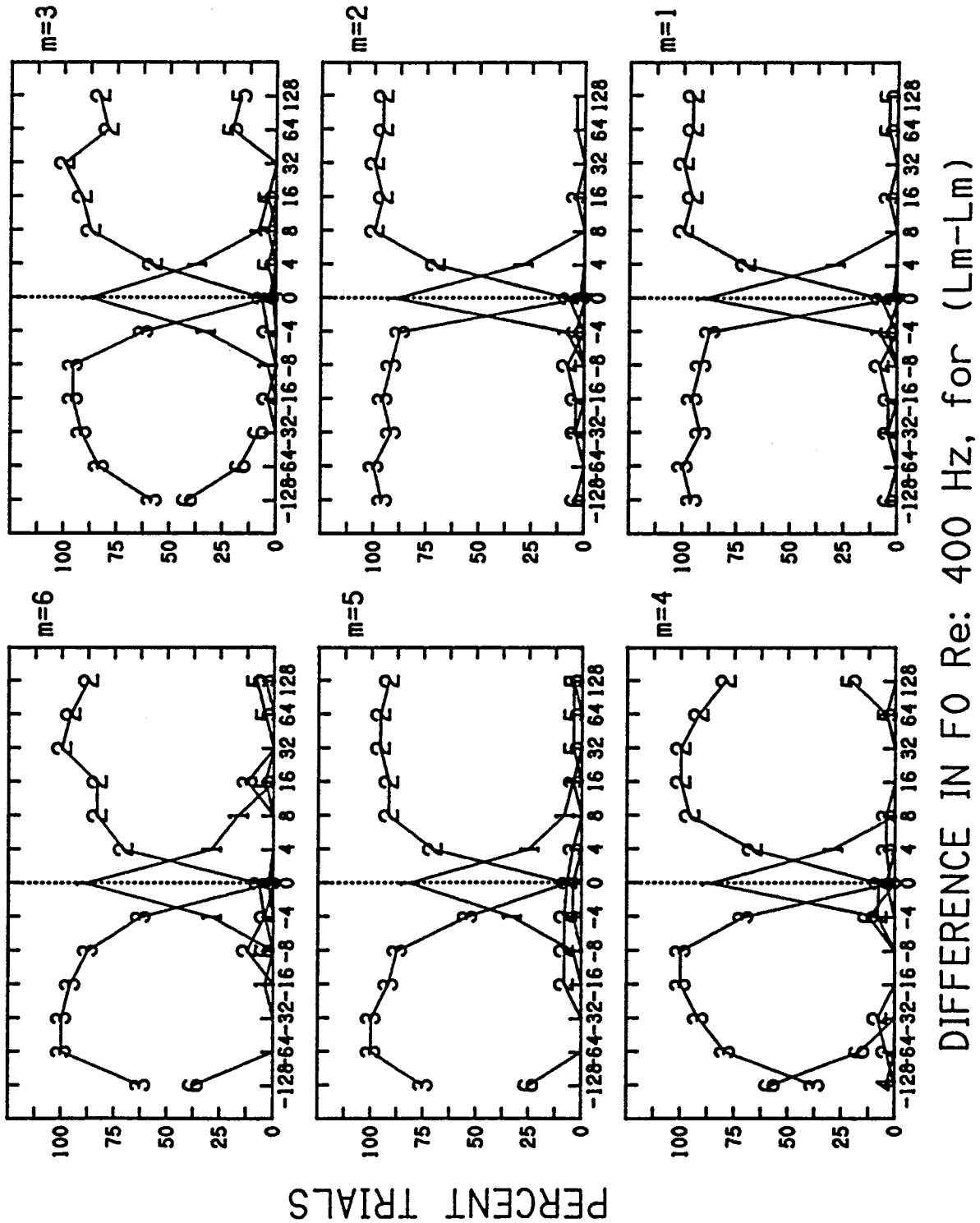
Figures 3.10 through 3.12 (next three pages)

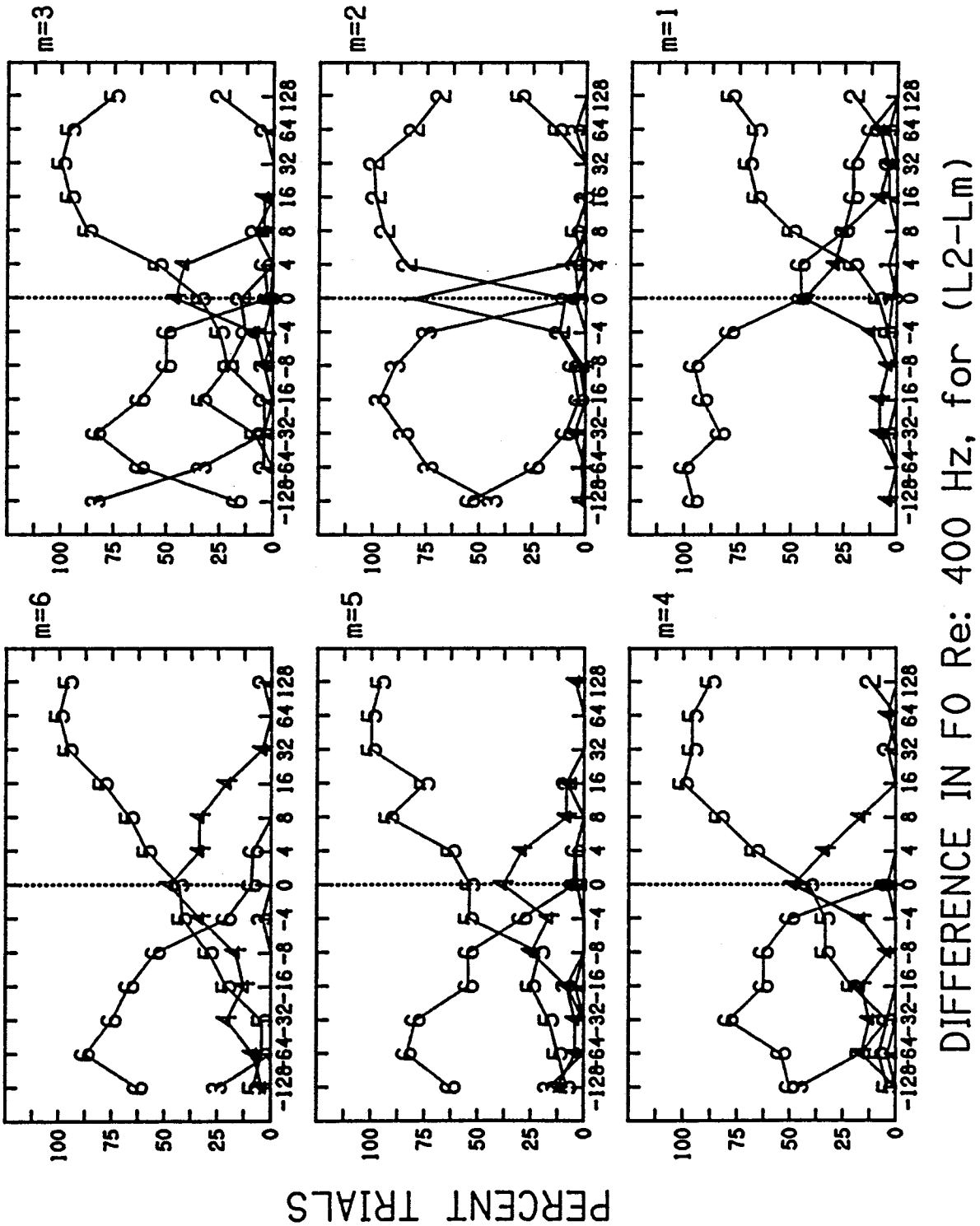
Each figure shows "magnitude-of-use" of different response labels averaged over 6 listeners, for locus relations between tones as mentioned at the bottom of the figure ((Lm-Lm), (L2-Lm) and (Lm-Lm), for figures 3.10, 3.11 and 3.12, respectively).

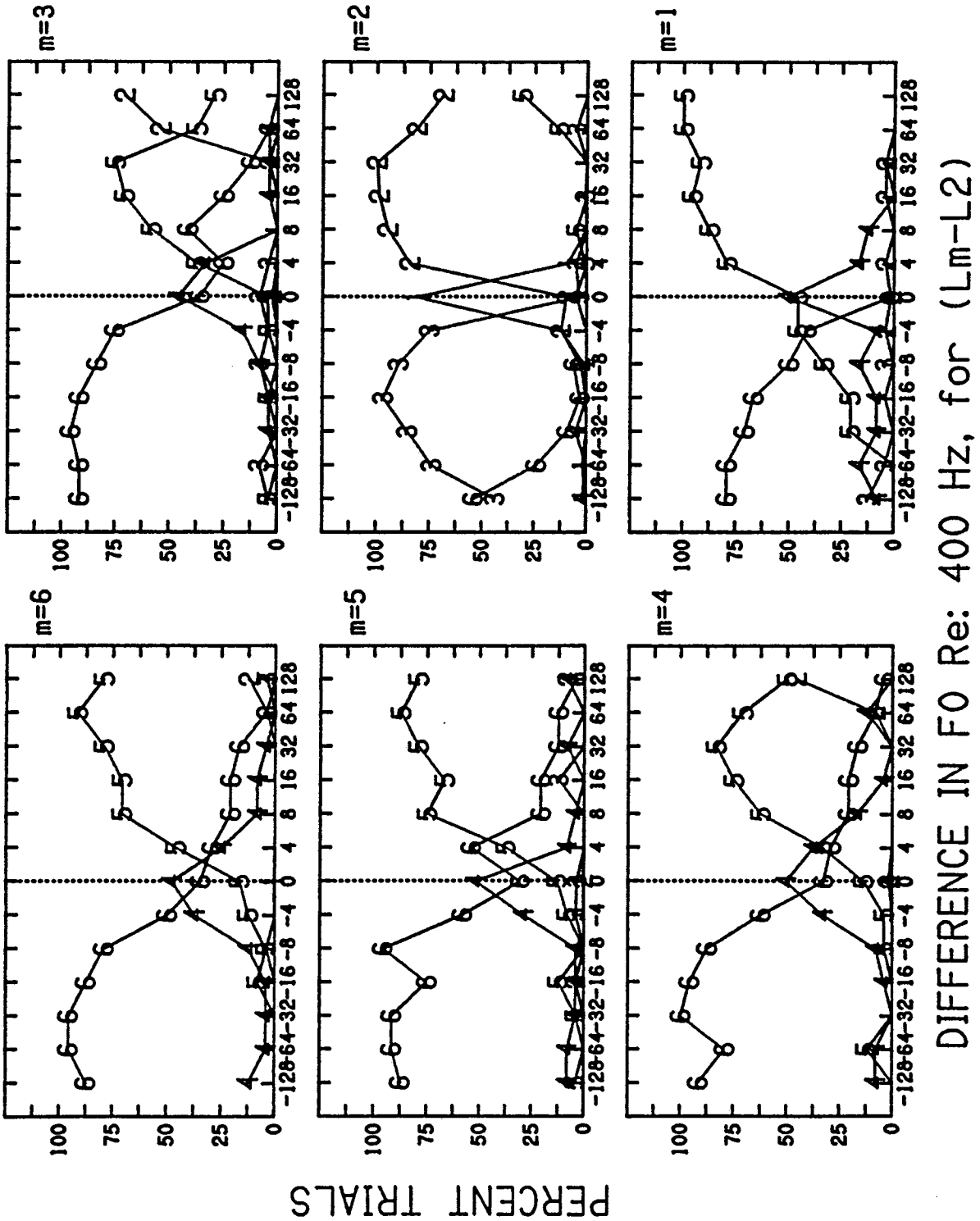
Each of the six "frames" corresponds to the value of the lower harmonic number "m" defining a locus. Differences in F0 between the two tones (re: 400 Hz) are displayed on the abscissa. The ordinate gives the proportion of trials on which different labels were used.

Numbers placed at data points correspond to the label number as defined in the task (illustrated below). Thus, '1' implies no change, '2', a rise in pitch without a change in timbre, '3' , a fall in pitch without a change in timbre, '4', a change in timbre without a change in pitch, '5' a change in timbre and a rise in pitch, and '6' a change in timbre and a fall in pitch.




m=3
m=2
m=1
m=6
m=5
m=4
PERCENT TRIALS
DIFFERENCE IN FO Re: 400 Hz, for (Lm-Lm)


m=3
m=2
m=1
m=6
m=5
m=4
PERCENT TRIALS
DIFFERENCE IN F0 Re: 400 Hz, for (L2-Lm)


m=3
m=2
m=1
m=6
m=5
m=4
PERCENT TRIALS
DIFFERENCE IN F0 Re: 400 Hz, for (Lm-L2)

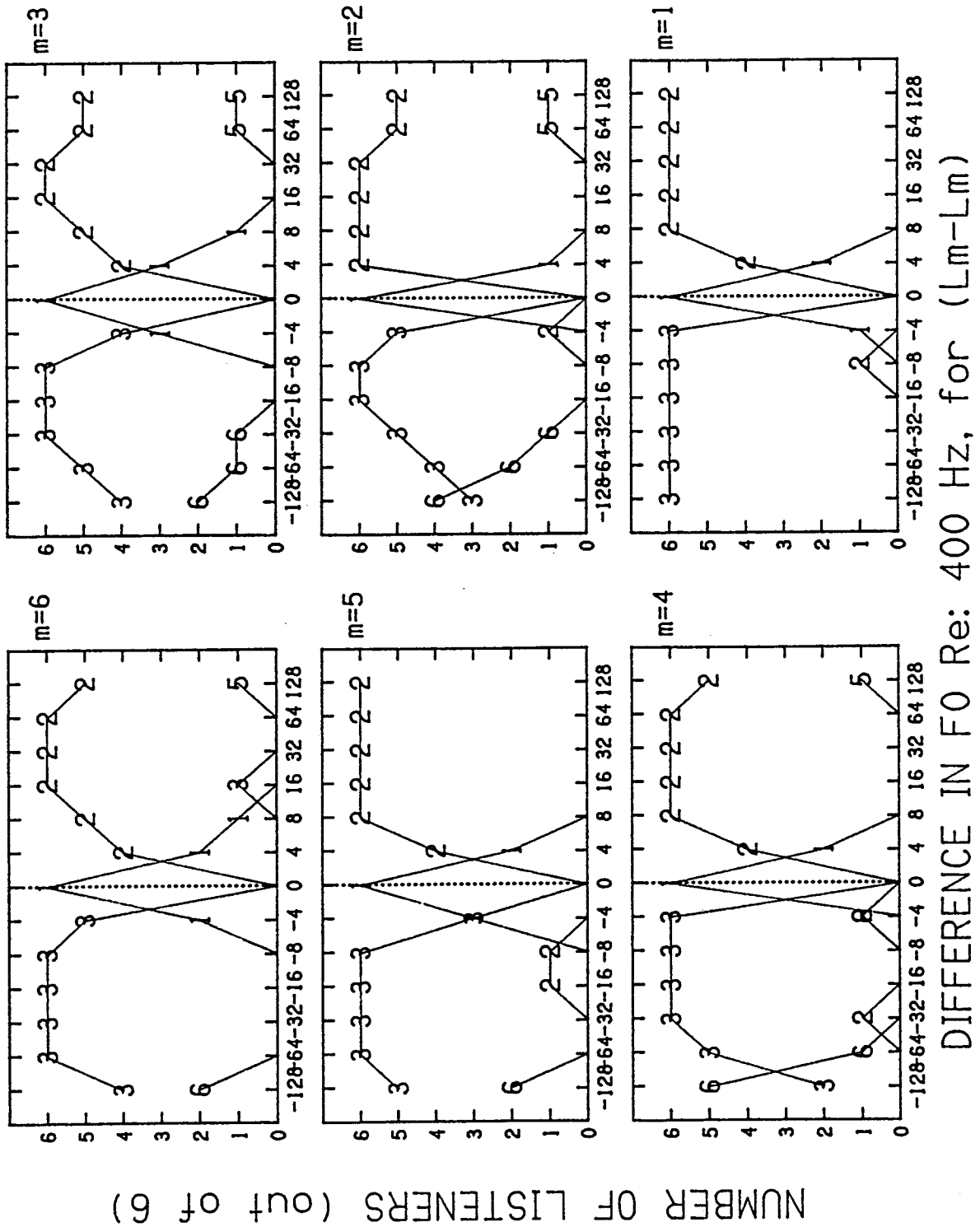
Figures 3.13 through 3.15 (next three pages)

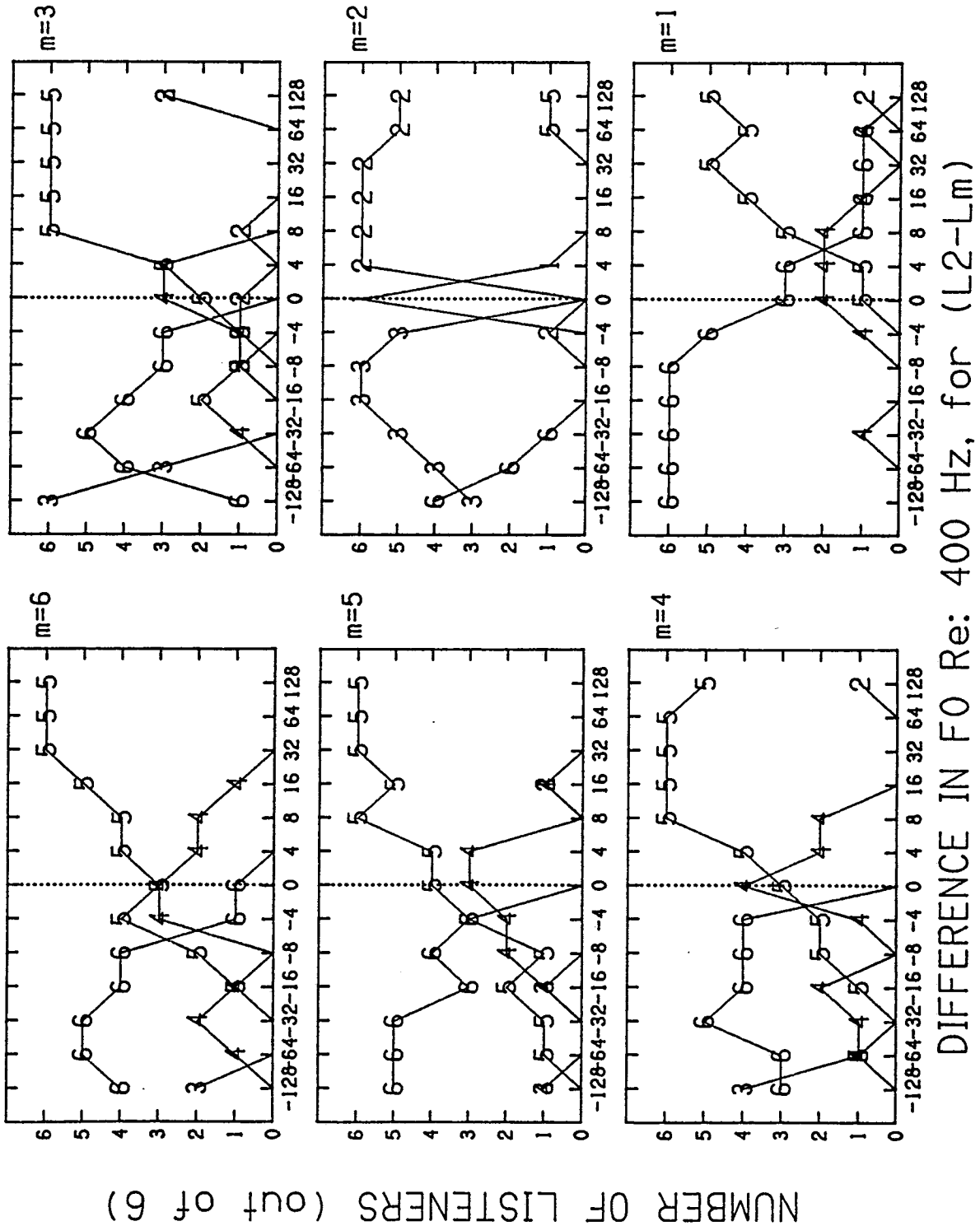
Each figure shows numbers of listeners and the "dominant label" selected by them for stimuli with locus relations between tones as mentioned at the bottom of the figure ((Lm-Lm), (L2-Lm) and (Lm-L2), for figures 3.13, 3.14 and 3.15, respectively).

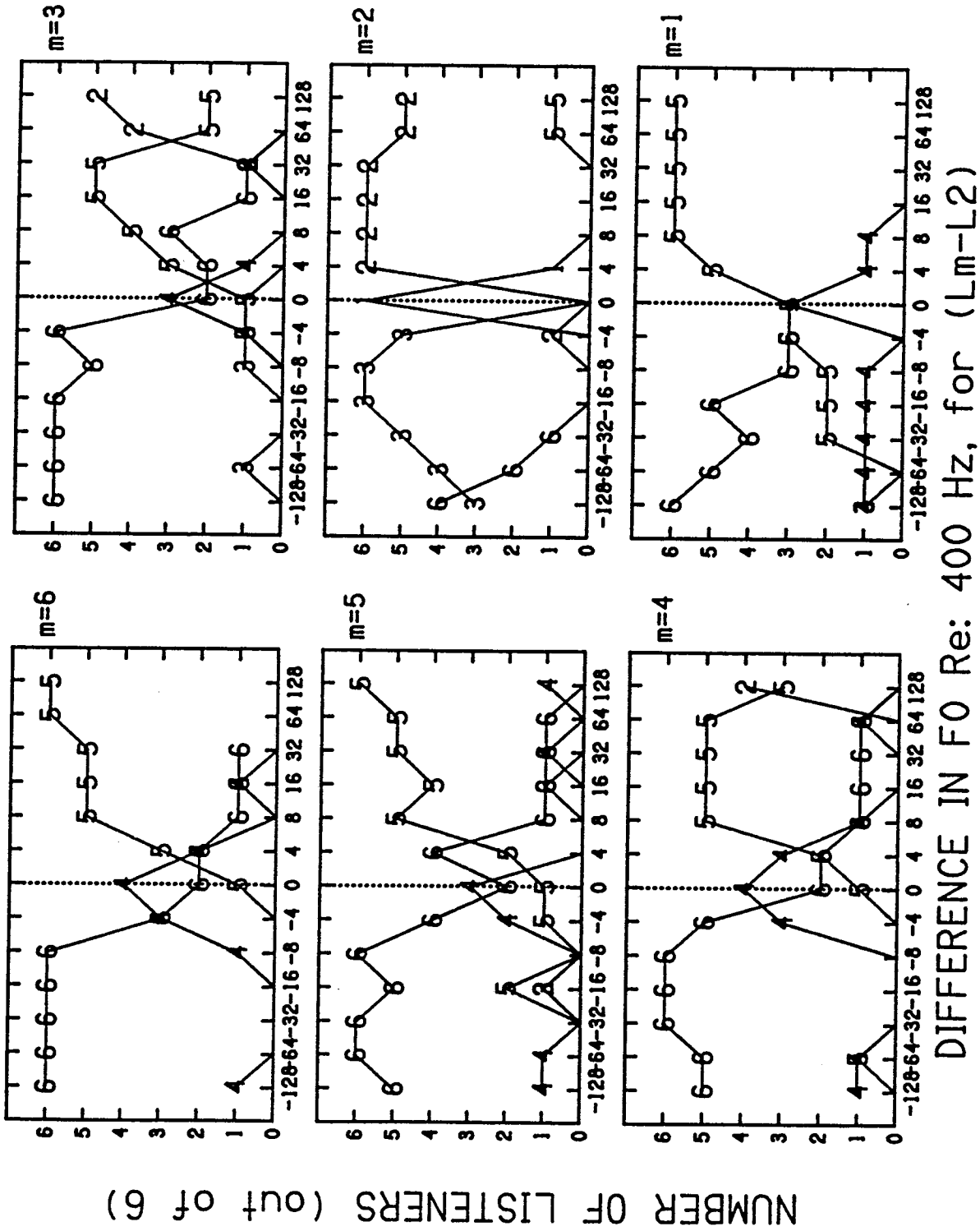
Each of the six "frames" corresponds to the value of the lower harmonic number "m" defining a locus. Differences in F0 between the two tones (re: 400 Hz) are displayed on the abscissa. The ordinate gives the number of listeners who selected the different labels that are plotted as data points.

As before, the data points correspond to the label number as defined in the task. Thus, '1' implies no change, '2' implies a rise in pitch without a change in timbre, '3' , a fall in pitch without a change in timbre, '4', a change in timbre without a change in pitch, '5' a change in timbre and a rise in pitch, and '6' a change in timbre and a fall in pitch.

Caution: Do not confuse the number of the response labels '1', '2', '3', '4', '5', '6' with the number of listeners (out of 6) represented along the ordinate.


m=3
m=2
m=1
m=6
m=5
m=4
-128-64-32-16-8-4 0 4 8 16 32 64 128
0 1 2 3 4 5 6
NUMBER OF LISTENERS (out of 6)
DIFFERENCE IN F0 Re: 400 Hz, for (Lm-Lm)


m=3
m=2
m=1
m=6
m=5
m=4
-128-64-32-16-8-4 0 4 8 16 32 64 128
0 1 2 3 4 5 6
NUMBER OF LISTENERS (out of 6)
DIFFERENCE IN F0 Re: 400 Hz, for (L2-Lm)


m=1
m=2
m=3
m=4
m=5
m=6
-128-64-32-16-8-4 0 4 8 16 32 64 128
0 1 2 3 4 5 6
DIFFERENCE IN F0 Re: 400 Hz, for (Lm-L2)
NUMBER OF LISTENERS (out of 6)

3.7 Discussion :

3.7.1 Influence of timbre on judgment of pitch change

The results of the present experiment confirm the relation between timbre and spectral locus, and further, illustrate the influence of spectral locus on perception of pitch.

Since the time of von Helmholtz, timbre has been considered to be related primarily to the steady-state Fourier spectrum described by the amplitude pattern of the various frequency components comprising a sound produced by a particular source. This physical spectrum is in turn related physiologically to the "region of maximal stimulation on the basilar membrane" (Ritsma, 1963). Recall that the region of stimulation on the basilar membrane is also a measure of "place" pitch (Licklider, 1954). This type of pitch corresponds to the "body" pitch described by Davis et al. (1951) and "spectral pitch" described by Terhardt (1974).

For complex tonal stimuli, the place pitch is usually **subservient** to the overall pitch of the composite stimulus. The overall pitch (corresponding to Schouten's "residue", Davis et al.'s "buzz", and Terhardt's "virtual" pitch) usually corresponds to the **periodicity** or F0 of the stimulus and is the primary variable used to make melodies in music.

As mentioned in chapter 2 (section 2.9), spectral pitch and timbre are intrinsically related, given their **mutual dependence on the place**

of stimulation on the basilar membrane. Since spectral pitch contributes to *virtual* pitch as well, some confusions between timbre and the overall, virtual pitch would also be expected.

Such confusions were observed indeed in the present experiment for some listeners. These listeners reported a simultaneous change in both pitch and timbre for some sequences, despite the absence of a change in fundamental frequency, or in some cases, pitch changes were reported in a direction *opposite* to that of the direction of change in F0. The direction of pitch change reported, corresponded instead, to the direction of change of the spectral *locus*. An ascent in locus for concurrent $\Delta F0 = 0$ to - 2% was construed as a rise in pitch. A descent in locus was similarly construed as a fall in pitch, despite the absence of a change, or an increment in F0 ($\leq$2%).

The following of direction of locus change implies that listeners were sensitive to "spectral pitch". For a small range of F0 change (0-2%), spectral pitch dominated over the perception of virtual pitch. For larger values of F0 change, the virtual pitch percept took over. In both cases, timbre differences were perceived in addition to differences in pitch. It would seem, then, that **spectral components play multiple roles in the perception of pitch and timbre:**

The spectral location of components is the major bearer of timbral identity of a sound (Grey, 1975). However, the components are also compared for pitch. If changes in the relative spacing (F0) of components are smaller than some limiting "threshold" value, changes in absolute

frequencies will also dominate the sensation of pitch. The magnitude of such a "threshold", however, may be different for different listeners.

In the present experiment, many individual differences were noticed. At least one listener (LC), was consistently able to follow changes in relative spacing (F0), and ignore change in absolute frequencies as indicating a pitch change. He did not ignore the changes in absolute frequencies as indicators of timbre change, however. Timbre differences were reported when absolute frequencies changed, even if no pitch change was reported. Pitch and timbre thus appeared to be separable for this listener.

Other listeners, however, showed confusions and difficulty in separating timbre and pitch when both these percepts were dependent on absolute frequencies of components (for $\Delta F0 \leq 0\text{-}2\%$). When the relative frequency cue became available, (for larger values of $\Delta F0$), they too were able to separate pitch from timbre. This range of confusion is fortunately within the limits of a musical "semitone", the smallest pitch step in the equal-tempered scale ($\Delta F0 \approx 6\%$).

Some details of the individual differences observed between listeners, and possible effects of prior listening experience and musical or psychophysical training are discussed in the next two sections.

3.7.2 Individual differences amongst listeners

While there was general agreement between listeners for sounds

in which there was no conflict between direction of locus change and F0 change, individual differences were noted for stimuli with divergent information. The spread of label choice across listeners was shown in figures 3.7, 3.8 and 3.9 for reference F0=200 Hz. *Individual* labelling data for sequences with locus relations (L2-Lm) and (Lm-L2) at the 200-Hz reference F0 are presented for some listeners in **figures 3.16 and 3.17**, to illustrate how they differed from each other, while maintaining consistency within themselves.

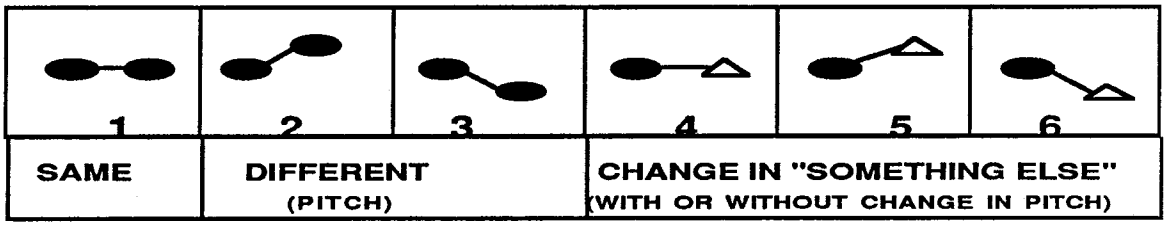
The "dominant" label choice for different stimuli is shown for listeners LC, GC and WK. These 3 listeners have been selected because of the apparent differences in their response strategies. WK and LC exhibited two extremes of response behavior. WK tended to report the direction of locus change as the direction of change in pitch, while LC followed the change in periodicity (F0) as an indicator of pitch change. The data for musician GC are shown for comparison with fellow- musician WK and are illustrative of a flexible response strategy representative of the modal response trend.

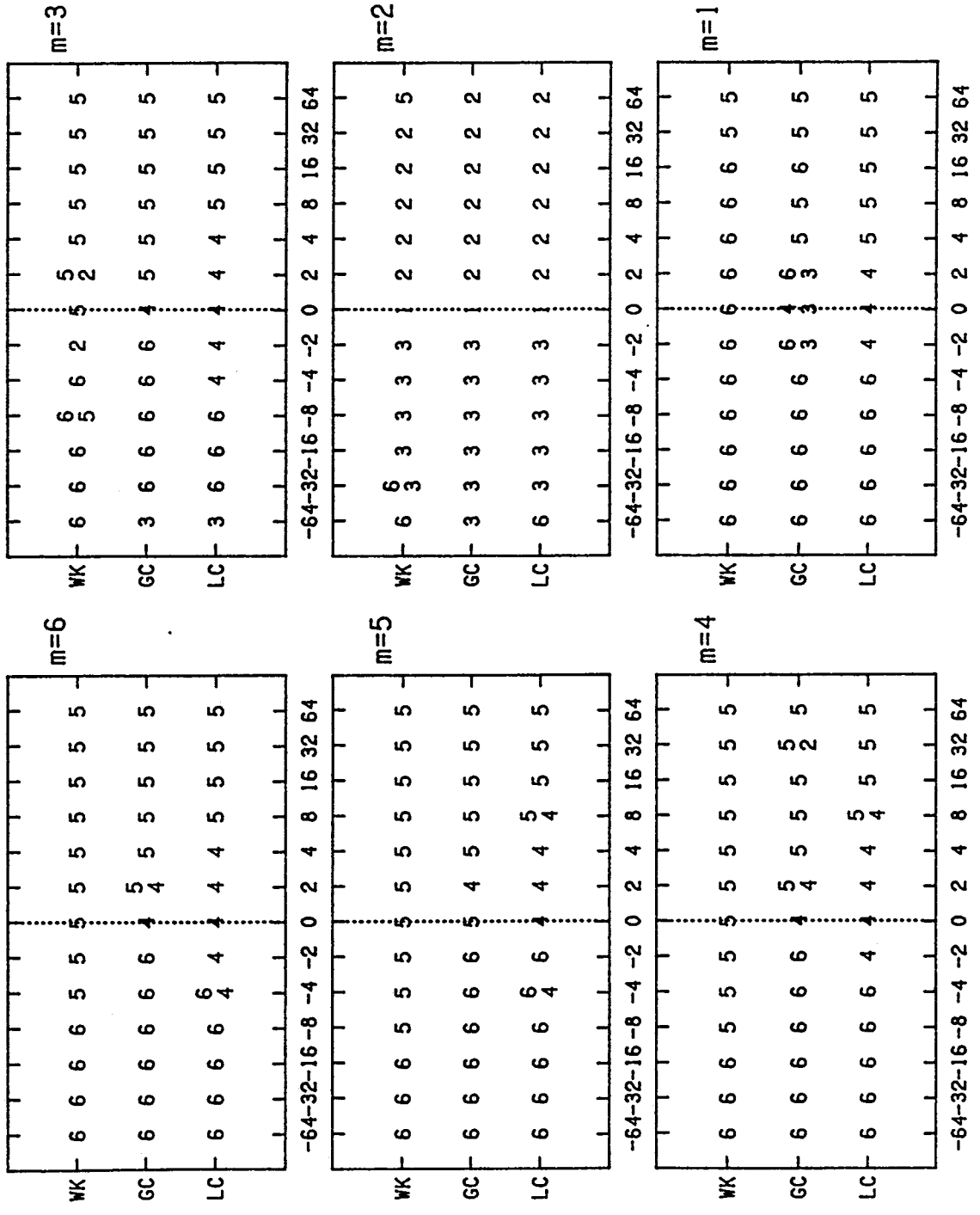
The general layout of figures 3.16 and 3.17 is similar to the previous graphs, with the difference that the three listeners are assigned different points along the ordinate.

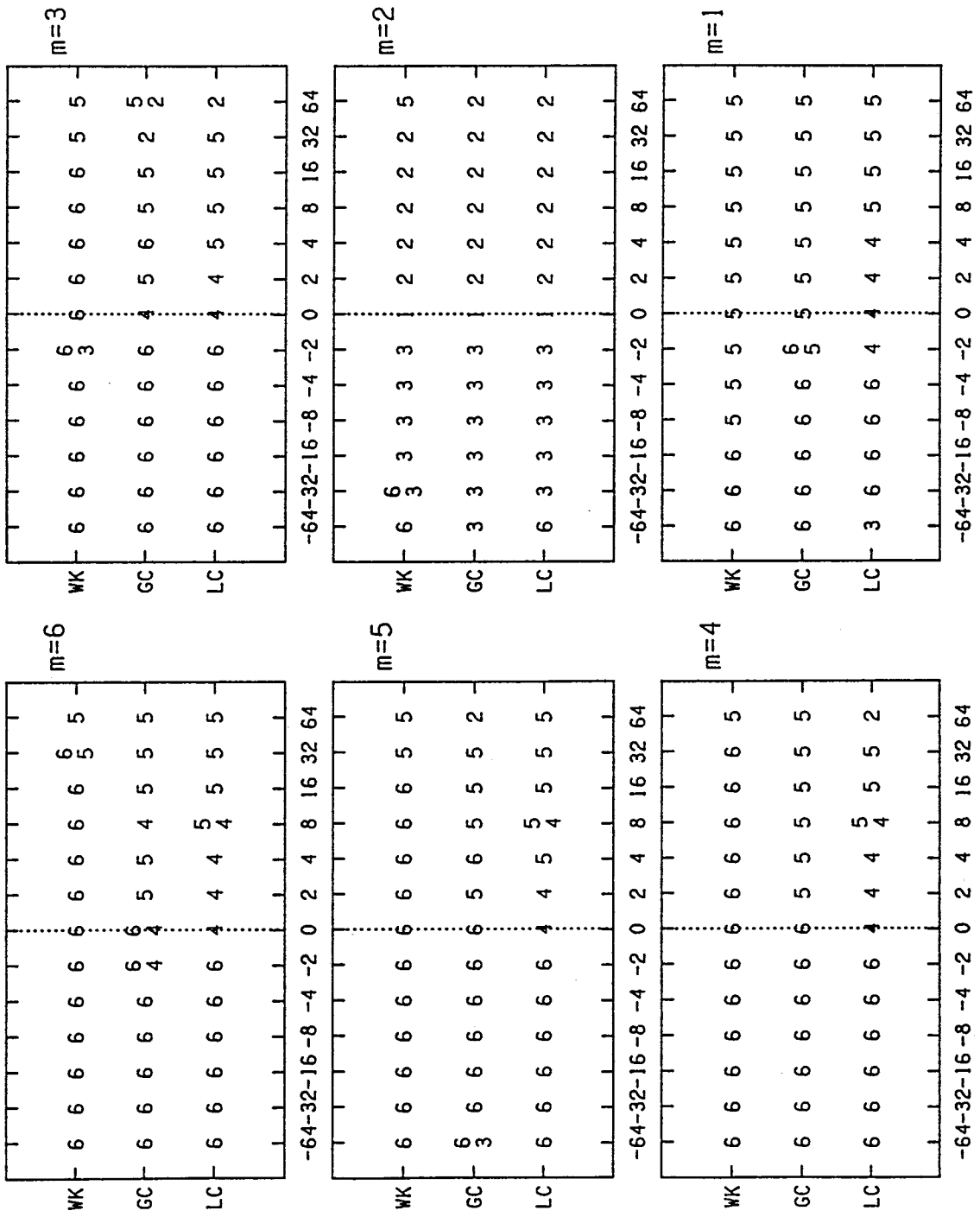
As can be seen in **figure 3.16,** for locus contrasts (L2-Lm, m>2), **listener WK** selected the response label '5' indicating a ***rise in pitch***, (along with a change in timbre), even for sequences where $\Delta F0=0$ and where F0 was *decreased* by as much as 8 Hz, (i.e., by -4%).

Figures 3.16 and 3.17 (next two pages) Individual listener data for sequences with locus relations (L2-Lm) and (Lm-L2), and changes in F0 (re:200 Hz) as shown along the abscisssa. Three listeners (LC, GC and WK) are assigned points along the ordinate.

The data points indicate the label selected most frequently by the listener in response to the different sequences. Label '1' implies no change, label '2' implies a perceived rise in pitch, label '3', a fall in pitch, label '4', a change in "something else", label '5', a rise in pitch and a change in "something else", label '6', a fall in pitch and a change in "something else" as is illustrated in the key below:




m=3
m=2
m=1
m=6
m=5
m=4
WK
GC
LC
-64 -32 -16 -8 -4 -2 0 2 4 8 16 32 64
INDIVIDUAL LISTENERS
DIFFERENCE IN FO Re: 200 Hz, for (L2-Lm)


m=3
m=2
m=1
m=6
m=5
m=4
WK
GC
LC
-64 -32 -16 -8 -4 -2 0 2 4 8 16 32 64
INDIVIDUAL LISTENERS
DIFFERENCE IN FO Re: 200 Hz, for (Lm-L2)

For sequence (L2-L1) with a descent in locus (bottom-right of figure 3.16), WK's label-of-choice changed correspondingly to '6', indicating a fall in pitch, despite an increase in F0.

A similar following of spectral locus is visible in **figure 3.17** for locus relations (Lm-L2). In fact, for sequences (Lm-L2, m>2) where there was a descent in locus, the paradoxical use of labels is even more prominent. WK selected the response label '6' indicating ***a fall in pitch*** even for sequences where $\Delta F0=0$ and where F0 was *increased* by as much as 32 Hz, (i.e., **by16%**). Sequences (L6-L2) and (L4-L2) (top and bottom frames on the left of figure 3.17) show this persistent following of the spectral locus, despite changes in F0. For sequence (L1-L2) with an ascent in locus (bottom- right of figure 3.17), the label-of-choice changed correspondingly to '5', indicating a rise in pitch, despite a decrease in F0, further verifying that WK was a "spectrum- follower", rather than a "period-follower" for values of period change $\leq 16\%$.

It could be suspected that perhaps the failure of WK to discriminate F0 is just a manifestation of a high DL for frequency. However, her judgments of pitch change for sequences with locus relations (Lm-Lm) indicated that she ***was*** able to discriminate changes in F0 as small as 1% and accurately identified the *direction* of change for all values of $\pm\Delta F0$. It thus seems that her choice of labels for sequences with locus contrasts (L2-Lm) and (Lm-L2) was not a consequence of any sensory limitations, but rather, reflected a criterion of judgment, based

on spectral locus.

In sharp contrast to the locus-biased performance of WK, **listener LC**'s ability to discriminate changes in F0 appeared to be *unhampered by concurrent changes in locus.* His responses indicate a tendency to follow "periodicity", and report pitch changes only when the periodicity (F0) of the stimulus was indeed changed. This is manifested by the use of label '4' at values of ΔF0 in the range bounded by ≤ ±4 Hz in figures 3.16 and 3.17.

The spread of label '4' around ΔF0=0 Hz seems to be a simple manifestation of a jnd-type limit in discriminating F0. The DL for frequency below 1000 Hz is indeed often accepted to be around 3 Hz (Durrant and Lovrinic, 1984).

Of the six listeners used, LC was unique in that he appeared to be largely unaffected by the absolute frequency of components in judgment of a change in relative spacing (i.e. F0) of the two "residue" sounds being compared. He *was,* however, affected by the absolute frequency of components in making judgments of timbre. Thus, the label '4', implying a change in timbre without a concurrent change in pitch, was used appropriately by him for sequences where locus changed but F0 did not.

Listener GC's choice of labels indicate a flexible response strategy. In the absence of changes in F0, he used label '4' (like LC) for most sequences with locus relations (L2-Lm, m≠5), as can be seen in figure 3.16.

For sequences with locus relations (Lm-L2), however, listener GC

responded more like WK, reporting direction of locus change as the direction of a perceived change in pitch. As can be seen in figure 3.17, he selected label '6' indicating a fall in pitch for sequences in which there was no change in F0, but there was a drop in locus. Similarly, he used label '5' indicating a rise in pitch for no change in F0 in the sequence with an ascending locus (L1-L2) .

Listener GC is a **highly trained musician**, as is WK. LC on the other hand, is not formally trained in music. The dichotomy in response strategy observed for WK and LC is very surprising, in light of this knowledge. Musicians are generally expected to be better followers of periodicity, while naive listeners are supposed to be more susceptible to being "deceived" by concurrent changes in spectral location (Dowling and Harwood, 1985; Krumhansl and Shepard, 1979).
Results of the present experiment are not in accord with this expectation.

Whether the divergence in response strategies is a function of different types of training, lack of experience, or based on more basic auditory limitations is difficult to infer. What *can* be inferred, however, is that **different listeners may be more sensitive to different aspects of the spectrum** (Lauter, 1985). Some listeners, like LC, may be more "tuned in" to a relative aspect of the spectrum related to F0. Others, like WK, may be more tuned in to the absolute spectrum. Still others, like GC, may choose to compare both relative and absolute relations between the spectra of stimulus tones.

Another inference that can be made from these data is that **all**

musicians are not alike. The difference in performance of musicians WK and GC, implies that even *within* the category of "musician", there exist individual differences. These differences may again be dependent on specific types of musical training (e.g. composition versus performance), or based on auditory factors peculiar to the listeners individually.

Listener GC has a Ph. D. degree in music, with an emphasis on composition. He also teaches "ear training and musicianship". His flexible performance is perhaps understandable from the viewpoint that musicians need to be able to listen both in an "analytic" mode that would be sensitive to the frequency composition of sounds, as well as in a "synthetic" mode that would be more sensitive to the larger sound (characterized by a unitary F0), rather than to its constituents.

Listener WK is a highly-trained performer on the double bass. Her inability to equate pitch based on equality of F0 despite differences in spectra is thus surprising. In performing with other musicians in an orchestral situation, F0 is equalized across instruments of different timbre. The timbre dimension related to spectral differences should thus be amenable to being "normalized" across pitch differences. However, this was not the case. The timbre difference between tones of different locus seemed to overshadow the true F0 difference for listener WK.

Her spectrum-biased performance may, however, be more understandable in light of the fact that she also composes from time to time, in the medium of "electronic music". Synthesizers and computers -

the "instruments" of electronic music, offer extensive control over the acoustic composition and juxtaposition of sounds. A wide range of timbres and diverse ways of grouping them are thus possible. WK's considerable experience with multi-timbral devices may then have been an operative factor in her timbre-dominated performance in the present experiment.

3.7.3 Influence of training

The observation of differences in performance of listeners and the cues they apparently attended to, raises the question whether these differences could be removed by training that constrains the listener to respond in a certain way. There is some evidence to indicate that training can, indeed, have a big effect on how listeners attend to stimulus properties.

A study by Cross and Lane (1963) showed that listeners could be trained to categorize sounds into groups based on increasing F0 or increasing center frequency (F1) of a formant-like filter. Listeners who were "taught" via feedback that correct identification of stimuli constituted their being ordered in terms of ascending F0, extrapolated this criterion to the test situation where stimuli were presented with mixed configurations of increasing F0 and increasing or decreasing F1. Listeners who were taught to identify the training stimuli based on decreasing order of F1, were able to classify the test stimuli in terms of

F1 only, disregarding changes in F0. Cross and Lane comment that the perceptual correlates associated with these stimulus changes could have acted as powerful cues in guiding their listeners' decisions. Thus, the group of listeners trained on the F0 criterion probably learned to respond in terms of perceived differences in pitch. The group of listeners trained on the F1 criterion most likely responded on the basis of differences in perceived timbre, ignoring any changes in pitch. In this particular situation therefore, the possible ambiguity between perceived pitch and timbre was constrained by the training criteria used, that in effect told the listeners what to listen for in the test stimuli.

While training can be a powerful tool in honing the responding skills of a listener to comply with some pre-determined scheme, it is not desirable in experimental situations where **the percept** is the main focus of interest rather than optimal performance on a task that has a right or wrong answer.

A task that forces listeners to judge which tone in a pair was "higher" in pitch, overlooks the fact that a pitch change may not be perceived at all, despite a change in the frequency of one or more spectral components. Percent correct measures for such a task are meaningless, since the experimenter cannot judge the direction of perceived pitch change simply on the basis of direction of frequency change (Pollack, 1978; Risset, 1971). On the other hand, discrimination tasks that ask for simple "same"/"different" judgements may be too global. Such tasks make no assumptions about the underlying perceptual

cues that guide discrimination and thus yield no information about the various percepts that may come into play for different magnitudes of spectral change.

In the present experiment, listeners were presented with an assortment of stimuli in "warm up" sessions to get them familiarized with the sounds and with the procedure prior to data collection. Many of these stimuli were selected to be blatantly representative of the 6 labelling options. However, listeners were not specifically trained to ignore or use different cues related to different stimulus dimensions, and the use of feedback was deliberately avoided. The aim of the experiment was to examine the listeners' responses, as being representative of salient perceptual features of the sounds that they were attending to. Individual differences were welcomed, rather than feared, as they highlighted different ways in which information was extracted and used for discrimination of spectral changes.

3.7.4 Role of corresponding harmonics

Just as there are differences between listeners in terms of the salience of different perceptual cues, there are also differences between the stimuli themselves, that enhance or diminish the availability of certain perceptual cues (Lauter, 1983). In the present experiment, large changes in F0 and large changes in locus would be examples of stimulus features that emphasize pitch and timbre differences, respectively,

although different listeners may vary in the degree to which they use these percepts as cues for discrimination.

One controversial stimulus feature that may affect the salience of different perceptual cues for discrimination, is the harmonic composition of the sounds being compared (i.e. the rank numbers of constituent harmonics). Several studies indicated in the past, that listeners sometimes match spectral pitches of components , rather than a global pitch for the complex as a unit (Davis et al., 1951; Risset, 1971, 1986; Smoorenburg, 1970). It is thus not clear if the difference limens reported for pitch in many experiments, pertained to the global or the spectral pitch (Ritsma, 1963; Faulkner, 1985) or perhaps, to both . . .

The present experiment provided stimuli that could differ both in harmonic composition and in F0. The "loci" contrasted in different stimulus pairs varied in their degree of spectral overlap. Some pairs (Lm-Lm) shared **all** "corresponding" harmonics, others (L2-Lm) and (Lm-L2, m≠2, 6) had **some** corresponding harmonics, and (L2-L6) and (L6-L2) had **none.**

The labelling data for different locus-F0 relations revealed that **listeners compared both locus and F0** in making judgments of pitch change. The F0 cue dominated the perception of pitch. Given changes in F0 > 2%, the direction of pitch change reported, agreed with the direction of F0 change. Within the range 0-2%, however, spectral pitches were compared by most listeners.

Furthermore, comparison of components was indicated for **all**

locus contrasts, not just those with "corresponding" harmonics. Sequences with loci (L2-L6), and vice versa, were also susceptible to the same spectral vs. virtual pitch dilemma for small values of ΔF0 despite the lack of "corresponding" harmonics. Thus, the direction of locus change (i.e. direction of component-frequency change) was reported as the direction of pitch change even for these sequences.

The claim of Faulkner (1985) that residue pitch is compared ***only*** for those sounds lacking common harmonics, is **not borne out** by the results of the present experiment. If this claim was valid, judgments of pitch should not have suffered for loci (L2-L6) and (L6-L2). Only a timbre change should have been reported for ΔF0=0 Hz, via the use of label '4'. Instead, labels '5' and '6' were used for the range ΔF0=0-2% by many listeners, indicating changes in perceived pitch in addition to changes in timbre.

As can be seen in summary table 3.5, the label '6' was chosen to indicate a fall in pitch for (L6-L2), with no change in F0, and for increments < +8 Hz re: 200 Hz (i.e. a 4% change). For the 8 Hz increment, the label '4' indicating *no change in pitch* was the "dominant" label. Again, this choice is indicative of comparisons of both relative and absolute frequency of components. The upward change in F0, counteracted by the descent in locus, seems to have evoked a balancing of pitch for the two sounds being compared. A timbre change, however, was still perceived.

The spectral frequencies thus appeared to play *at least two roles* in

discrimination. For the most part, they are the bearers of timbre. To some limited extent, they also influence pitch.

The preoccupation of many experiments with the global, "musical" pitch of complex tones related to F0, has led to oversight of other perceptual roles played by the components actually present in the spectrum. It seems that these spectral components can do more than simply contribute their frequency values for calculation of a common factor to be interpreted as the "pitch" (Goldstein, 1973). Rather, they can make their presence felt in terms of **both timbre and pitch differences.**

3.7.5 Possible implication of results for "pitch shift" experiments

The results of the present experiment indicate an influence of locus on judgments of pitch. A similar "pitch shift" effect was also reported in earlier experiments on pitch of inharmonic complexes (de Boer, 1956/1976; Schouten et al., 1962).

The present experiment employed harmonic complexes with or without the fundamental, while the "pitch shift" experiments typically used residue tones, that were made *inharmonic* by a linear shift in the frequencies of all the components. The two types of experiments are thus not totally equivalent. However, the dependence of perceived pitch shift on absolute frequency noted in the present experiment, is similar to

the "first effect" of pitch shift described in **section 2.8.7,** that showed a dependence of the pitch shift (Δp) on the deviation in frequency (Δf) of the central component 'n' in a series of components linearly shifted in frequency from harmonic values ($\Delta p = \Delta f/n$).

In the present experiment, the role of spectral locus as contributory to both spectral pitch *and timbre* is acknowledged and anticipated in the design of the task. In de Boer's pitch shift experiments, the percept of interest was the low, "virtual" pitch of the complex. Possible timbral changes ensuing from the changed frequencies of components were not mentioned. A harmonic residue signal was used as the comparison signal who's pitch was to be matched with that of the inharmonic residue. Details of the spectral composition of the matching signal are not given in the description by de Boer (1976). The results of the present experiment indicate that this may be a crucial factor in making pitch comparisons. The pitch shifts obtained by de Boer could well be a consequence of attempts at equalizing the pitch difference that would be perceived *a priori* if the two signals varied substantially in their spectral composition.

These suggestions are intended only to raise consciousness about the multiple percepts associated with spectral changes in signals, and are not intended to criticize the earlier work on pitch as being negated. On the contrary, the earlier work offers a rich store of data and stimulus situations that can be re-examined for insights about the perceptual cues guiding performance.

3.7.6 Spectral locus, "sharpness", and "tone height"

The absolute frequencies of components contribute to both the timbre and the pitch of complex sounds. The terms "sharpness" or "brightness" have been used to describe timbre differences related to the "spectral center of gravity" of the stimulus (von Bismarck, 1974; Wessel et al., 1987; see **section 2.7**). The term "tone height" has been used to describe a dimension of pitch related to location of the spectrum (Shepard, 1982).

The structural model for pitch in terms of a helix comprising a rectilinear component related to "tone height" and a circular component related to "tone chroma" was described in **section 2.8.6.** In physical terms, the "tone height" dimension is related to the absolute frequency of the stimulus while "chroma" is related to the fundamental frequency via a multiplicative relation involving octavial ratios.

Given their mutual dependence on absolute frequency, these "sub-dimensions" of pitch and timbre should exhibit correlated behavior when frequencies of components are changed, i.e. a "sharper" timbre should be construed as having greater "tone height".

Patterson (1989 a, b) has recently begun investigations into the perception of "tone height" for multi-harmonic complex tones constrained to have the **same chroma** (equal F0) along the circular dimension of the pitch helix. Some of his preliminary experiments indicate that spectral filtering (amongst other variations) can have a profound impact on the

perception of tone height. If the low components (1-7) are removed from a 24-component complex tone, listeners judge the tone height as being higher (manifested by choice of a higher octave representing the perceived location of a "test" note). This difference in perceived tone height is claimed by Patterson (1989 b) to be a *"substantial component of many presumed timbre differences"* for stimuli with the same chroma, but different spectral compositions.

A similar influence of spectral composition on judgments of the octave position of notes was also reported by Hesse (1982). Given a series of notes of differing "brightness" (related to high-frequency content), listeners reported the "brighter" notes as belonging to a higher octave. Hesse thus concluded that "the octave position to which a listener estimates a note belongs, is not determined merely by the fundamental frequency - a leading role is also played by the brightness of the sound, the latter being dependent upon the structure of the spectrum" (p. 219).

In the present experiment, many situations were observed where a change in locus *without a change in F0* was interpreted as a change in pitch in the direction of locus change. In terms of the helical model, the tones had the same assigned "chroma" (F0). Given the equality of F0, the two tones of such sequences should not have evoked changes in tone height related to octave differences. However, changes in pitch *were* reported, as were changes in timbre.

In line with the observations of Hesse and Patterson, it seems that "sharper" or "brighter" timbres, correlated with loci in higher frequency

regions , were construed as having a higher pitch. The design of the present task did not enable measurement of perceived *magnitude* of pitch difference, so it cannot directly be inferred if these reports of pitch change were "octave" or "register" errors as found by Hesse. The inference is limited to the *direction* of perceived pitch change, which was correlated with the direction of locus change.

The helical model for pitch appears inadequate to deal with situations of this type, in which F0 (and thus chroma) are equated, the relative spacing (or F0) of components is *not related by an octave* (doubling factor), and yet tone height is perceived to increase with sharpness of timbre.

3.7.7 Competition between spectral locus and F0

In light of the locus-dominance effects described in the last section, one could arrive at the simple conclusion that stimuli comprising high components will be judged as being "higher" than stimuli comprising low components, despite equivalence of fundamental frequency ! While this may be the case indeed, for tones of equal F0, evidence also exists to show that *changes in F0* interact and compete with *changes in locus* to indicate direction of pitch change.

In the present experiment, spectral locus continued to play a dominant role over F0, for concurrent values of F0 change ≈2% (4% for some listeners). For larger values of ΔF0, however, F0 became the

dominant cue for pitch. Beyond the interaction range (extending from equal F0 to the 2%-4% limit), direction of reported pitch changes became congruent with the direction of F0 change despite large jumps in spectral locus. For some listeners (e.g. LC), the F0 cue was dominant in all cases, as the bearer of pitch.

The interesting experiments of Smoorenburg (1970) and Risset (1971, 1986) also indicated interactions between locus and F0, and differences in the ability of listeners to use these variables as indicators of pitch change (see section 2.9 for details).

In Risset's experiment, a tone B', with F0 =2 X the F0 of tone A', was judged to be *lower,* rather than an octave higher in pitch than A', by almost all listeners (50/53). Scrutiny of the stimuli revealed that components proximal in frequency were compared to yield the divergent judgment of pitch.

This situation highlights a **paradox** in the perception of pitch: Both the F0 and the locus (components) of tone B' were "higher" than those of tone A'. There should thus have been no conflict in judgment of pitch change. The only apparent reason for judging A' to be higher, appears to be the fact that the $n+1^{th}$ harmonics of A' were higher in frequency than the n^{th} harmonics of B'. Instead of comparing "corresponding" harmonics *a la* Faulkner (1985), listeners appear to have compared harmonics that were *spectrally close in frequency.*

This observation is of relevance to the issue of corresponding harmonics discussed in section 3.7.4. Rather than the equivalence of

harmonic numbers per se, it seems that it is the **degree of spectral overlap** that plays a role in determining if comparisons of component frequencies will yield changes in pitch. If there is no, or little, spectral overlap between components of adjacent tones, comparison of frequency of *individual* components may be superseded by perceived changes in timbre related to the gross difference in spectral envelope (or locus). However, these timbre changes may in turn impair judgment of F0 change, and induce reports of pitch jumps related to the change in gross frequency region.

If two tones presented for pitch comparison do not have any overlapping harmonics, this *does not imply* that component frequency comparisons have been prevented. All it means is that the components being compared are even further apart in frequency! The jump in locus will be correlated with a change in timbre, which in turn will impede the accurate judgment of "residue" pitch. The higher DLs obtained for residue tones may thus indeed be a consequence of timbre interference, as suggested by Moore (1987), rather than reflecting a comparison of "true" pitch as claimed by Faulkner (1985).

The present experiment, as well as the results of Smoorenburg and Risset validate the perception of such changes in timbre, and consequent confusions in pitch judgments.

3.7.8 Pitch and timbre tradeoffs

An earlier experiment by the author (Singh, 1987; described in chapter 1), also revealed interaction between pitch and timbre based on relative differences in spectral locus and spectral spacing (F0).

Stimuli very similar to those of the present experiment were used in a **"streaming"** paradigm. Two adjacent tones (say, A-A*) of equal F0 but different loci, forming part of a larger 4-note sequence (A-A*-B-B*), were perceptually segregated from each other and grouped with alternate tones of different F0, that were nearer in absolute frequency region and perceived to be of the same timbre (A-B, A*-B*).

This dominance of timbre as a cue for segregation over pitch, persisted until the F0 ratio of the two pairs of tones exceeded 3:2 (a "musical fifth"). Pitch then took over as the dominant cue guiding perceptual grouping (yielding the groups A-A* and B-B*). Interplay between absolute and relative aspects of the spectrum, quantified by spectral locus and spectral spacing, again seems to be the underlying basis for this pitch-timbre tradeoff.

3.8 Conclusions:

Based on the results of experiment 1, three spectral features emerge as important factors in the perception of changes in pitch and timbre of sequentially presented harmonic complex tones:

1) Spectral spacing or F0,
2) Spectral locus, and
3) Spectral context.

It appears that listeners perceive pitch and timbre changes in the **context of the overall distribution of spectral energy** of complex tones. Rather than the comparison of "corresponding" or "coincident" harmonics, **a general principle of spectral matching** is implied. The matching procedure appears to comprise two operations: one related to comparisons of relative frequency relations (such as F0 and harmonic "number"), and the other related to comparisons of the absolute frequency of components (such as locus contrast).

Spectral spacing or F0 is the dominant bearer of the pitch of a complex tone (the "chroma"). However, small changes in F0 may be overshadowed by larger changes in spectral frequencies of components. These changes, characterized by spectral "locus", are primarily perceived as changes in timbre, but also contribute to the perception of pitch. The frequency information obtained from spectral components is usually integrated to yield an overall frequency value related to the perceived pitch of a complex tone. For harmonic complex tones, this value typically equals the fundamental frequency (F0).

In some situations however, the spectral distributions of adjacent tones seem to be compared at a primary level (before central pitch processes take over the pitch-derivation operation). S pectral pitches of

individual components may be perceived at this primary level, along with a gross measure of the spectral envelope characterizing the timbre.

For the type of spectral loci employed in the present experiment, timbre changes appear to be related to the "sharpness" dimension described by von Bismarck (1974 b). Contribution of "place" cues to both spectral pitch and timbre results in the interaction observed between these percepts. An ascent in spectral locus yields a sharper timbre. Since "sharpness" too can be scaled on an ordinal scale from low to high, this timbre change may be construed as a pitch change.

If changes in locus and F0 are in the same direction, there is no conflict in judgments of direction of pitch change. If they are in opposite directions, the direction of the locus change may be reported as the direction of pitch change, despite changes in F0 in the opposite direction. This confusion prevails over a modest range of ΔF0, from 0-2%. For larger ΔF0, the periodicity of the stimulus takes over as the primary cue for pitch, and timbre and pitch become separable.

CHAPTER 4

EXPERIMENT 2

Perceptual correlates of changes in frequency for harmonic complex tones and inharmonic complexes with equal spacing between spectral components

4.1 Introduction :

The stimuli employed in experiment 1 comprised harmonic complex tones that were made to differ in spectral locus and/or spectral spacing. Since the components of each complex were successive *harmonics,* spectral spacing was equal to the fundamental frequency (F0). The results of experiment 1 indicated that judgments about pitch change were not simply dependent on F0 change (or relative spacing change), but could also be influenced by the difference in spectral regions of the tones being compared.

This influence of absolute frequency position of components on the residue pitch of a complex as a whole is reminiscent of the "first effect" of pitch shift described previously in Chapter 2 **(section 2.8.7).** This effect, reported first by Schouten (1940), then studied extensively by de Boer (1956/1976) and Schouten et al. (1962), was noticed in experiments using inharmonic signals with regularly-spaced components. Such signals were usually generated by amplitude modulation of a carrier

frequency 'f' by a modulating signal comprising components with frequency g, 2g, 3g etc. By adding the carrier f to the spectrum resulting from modulation, a complex signal could be generated with components:

f-3g, f-2g, f-g, f, f+g, f+2g, f+3g etc.

If **f=ng**, such a complex is **harmonic**, with all components being integral multiples of g as well as being spaced by a frequency difference (= 'g') in the spectrum. However, f *need not be a multiple* of g. In that case, the components of the resultant spectrum are still spaced by g, but the complex becomes inharmonic.

The main thrust of the early experiments using signals of this type was to demonstrate that a change in the "low", residue-type pitch was perceived when components of a harmonic complex were shifted linearly in frequency. This shift in pitch implied that the residue pitch did not simply depend on the period of the resultant waveform, but took into account "place" factors related to the absolute frequencies of components as well.

It was suggested by de Boer (1976), that both temporal and spectral comparisons could be made in pitch estimation. Comparisons of temporal fine structure could be made to yield a "pseudo period" while the frequencies of components could be compared to yield a "pseudo fundamental" that provided the best fit to the spectral pattern The low pitch could then be derived from the inverse of the pseudo-period or from the value of the best-fitting pseudo- fundamental.

The pitch shift observed for such stimuli was found to be primarily dependent on the carrier frequency via the simple relation: $\Delta p = \Delta f/n$, where Δp refers to the frequency equivalent to the pitch of the signal, Δf refers to the change in carrier frequency (or center frequency of the distribution) and n is an integer corresponding to the rank order of the central component. The pitch of the complex usually changed in the direction of change of the center frequency and was predictable from the linear relation given above.

However, in some situations, the pitch shift was observed to deviate from this simple proportional relation. The deviation from proportionality was termed the "second effect" of pitch shift. A manifestation of this effect was the observation that pitch went down (rather than up), when f was kept constant but the spacing g was increased.

The deviation from a simple linear relation between Δf and Δp can be described empirically by the formula (from Schouten et al., 1962):

$$\Delta p = [(1+b)/n]\Delta f - b\Delta g$$

where "b" is an empirical constant that is a measure of the second effect. Thus if b=0, the linear relation is obtained. If b is non-zero, the pitch shift will depend both on the change in center frequency, and on the change in spacing.

The center frequency 'f' for modulated signals of this type can be considered to be representative of the spectral "locus" of the sound, since

the components are spaced symmetrically around this mid- frequency. The influence of locus on perception of pitch has already been shown in the first experiment. In order to further explore the **effect of spacing**, stimuli for a second experiment were designed to resemble the type of inharmonic signals employed in the pitch-shift studies.

However, the stimuli were not typical "residues" lacking the fundamental. Instead, a fairly dense spectrum comprising 10 harmonics of equal amplitude was used as the reference sound. A following comparison sound was made to differ either in terms of F0, or in terms of deviations in frequency of components away from standard "harmonic" values. The F0 changes preserved harmonicity, but resulted in a change in spectral spacing, while the linear changes were inharmonic and resulted in constant spacing, but a small shift in absolute frequencies of components. The aim of the experiment, as before, was to ascertain the perceptual correlates of the spectral changes.

In addition to the expected pitch changes, it was anticipated that **timbre changes** may ensue as well because of the loss of harmonicity. While the preservation of spacing preserves the period of the temporal envelope, deviations in frequency away from harmonic values lead to a drifting of the phase relations between components. The fine structure of the waveform within the gross lobes of the envelope thus changes (de Boer, 1976).

According to a "phase rule" proposed by de Boer (1976), "the

phases ϕ_i of the components of a sound can be altered by an amount $\Delta\phi_i$ that is a linear function of frequency without causing a change of timbre" (p. 518). For small changes in frequency of components, this rule would predict that phase changes will not be noticed in terms of a timbre difference. Evidence for the validity of this rule is not generally available.

Rather, changes in waveform have been empirically observed to be correlated with perceived changes in timbre, often described as "roughness" or "dissonance" (Goldstein, 1967; Mathes and Miller, 1947; Plomp, 1970; Plomp and Levelt, 1965; Terhardt, 1974). It was thus expected that shifts in Δf that led to marked changes in the waveform could potentially yield timbre cues. A labelling task (similar to that used in experiment 1) was used with the hope that the extended response options would permit reports of timbre change, and enable exploration of the influence of different values of Δf on timbre.

Inharmonic stimuli are also associated with a **lack of fusion** (Cohen, 1980, McAdams, 1984). Since the time of von Helmholtz (1877/1954), it has been general knowledge that the ability to group partials into a single fused entity with a characteristic pitch and timbre is aided by the fact that the partials generated by a periodic sound source exhibit correlated behavior. Factors facilitating the perceptual fusion of components of complex sounds were discussed at great length in chapter 2. Harmonicity was stated to be one such factor that lends coherence to a

complex and aids in its perceptual fusion. Inharmonicity, on the other hand, opposes fusion, and partials sound disconnected. A second labelling task was thus designed to enable reporting of changes in the perceived fusion of the stimuli used.

4.2 Stimuli :

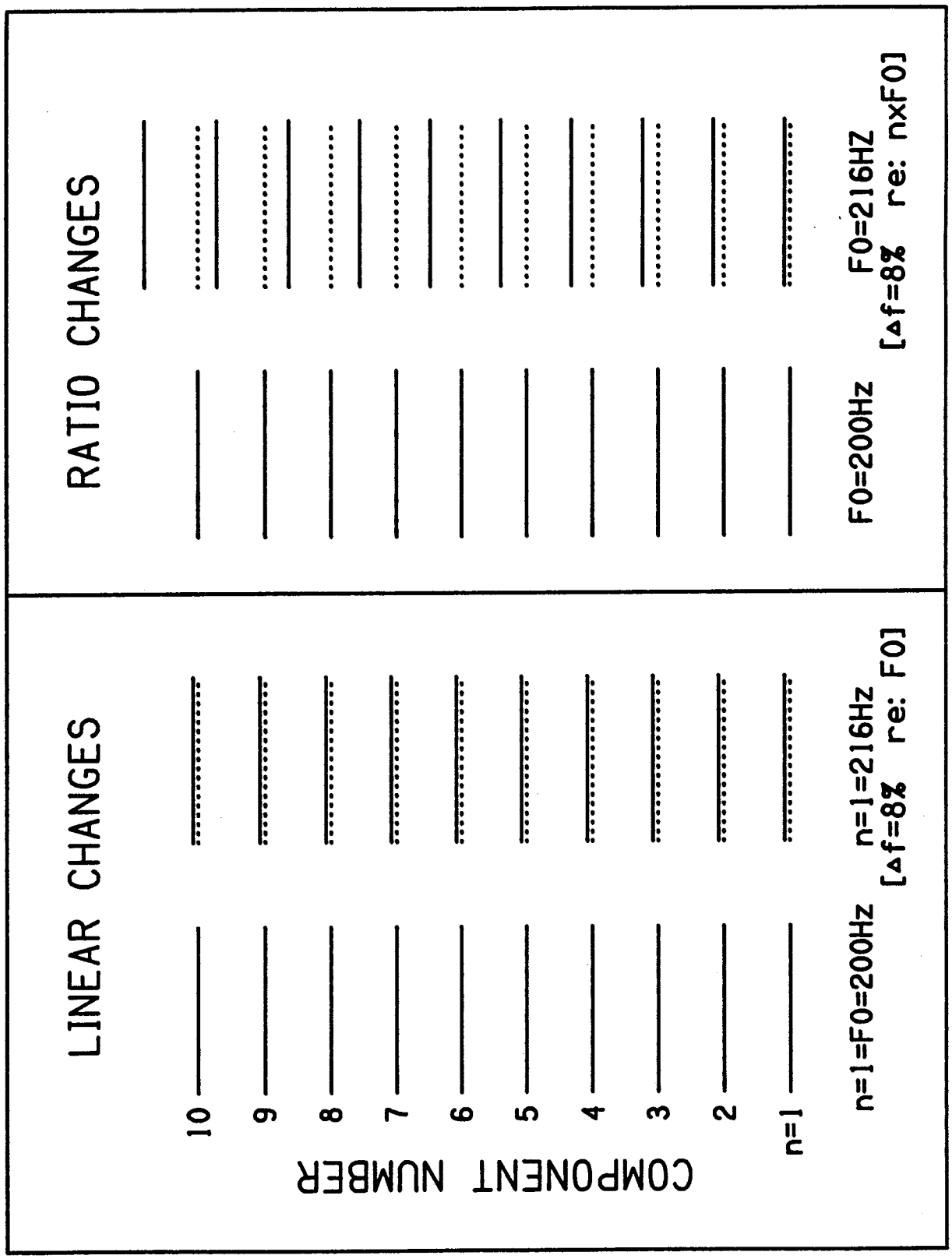
Two-tone sequences were used as stimuli, with the first tone set to be a "standard" complex with the first 10 harmonics of F0=200 Hz, or F0=400 Hz at equal amplitudes. The second tone was either identical to the first, or differed in that it had either all components displaced by the same frequency stepsize (linear shift), or by the same ratio (different magnitude of shift for different components). The **linear** steps used for frequency changes were $\pm$ 2, 4, 8, 16, 32 and 64 Hz for the 200-Hz standard, and $\pm$ 4, 8, 16, 32, 64, and 128 Hz for the 400-Hz standard. The **ratio** changes used were $\pm$ 1%, 2%, 4%, 8%, 16% and 32%.

All tones were 100 msec in duration and a 300-msec silent interval separated the two tones in a sequence. The tones were shaped with a cosine pedestal envelope with10-msec rise and fall times. The sequence was presented twice per trial, separated by a time interval of 1600 msec between repetitions.

Figure 4.1 illustrates the difference between spectra of stimuli with linear and ratio changes in frequencies of components.

Figure 4.1 (next page) Schematic representation of spectra of the type of stimuli used in experiment 2. The vertical dimension represents frequency, and the horizontal dimension represents time (not to scale). Each tone comprised 10 components.

The left panel shows inharmonic stimuli with linear changes of the same absolute magnitude (e.g. $\Delta f = 16$ Hz) made in the frequency of all components, while the right panel shows harmonic stimuli with all components of the second tone displaced by the same relative proportion (8%). Dotted lines represent original frequencies relative to which changes were made. As the figure illustrates, the lowest component in both cases undergoes the same magnitude of frequency change, but higher harmonics have greater deviations in the harmonic case.


LINEAR CHANGES
RATIO CHANGES
COMPONENT NUMBER
10
9
8
7
6
5
4
3
2
n=1
n=1=F0=200Hz
n=1=216Hz
[Δf=8% re: F0]
F0=200Hz
F0=216Hz
[Δf=8% re: nxF0]

Two pairs of sounds are shown in a spectrographic format in figure 4.1, with the vertical dimension representing frequency and the horizontal representing time (not to scale). The first sound in both pairs is a standard harmonic complex, while the second sound is harmonic (in the pair on the right), or inharmonic (in the pair on the left), depending on the type of change made in the frequency of components. The dotted lines in the second tone represents original frequencies, while the solid lines represent changed frequencies.

The figure shows that linear changes preserve the spacing of components, with all components being deviated by the same absolute amount in frequency with respect to the corresponding components of the standard. Ratio changes on the other hand, lead to a change in the spectral spacing. The components are deviated by the same relative proportion with respect to their counterparts in the standard tone, but differ in the absolute amount of frequency shift in Hz.

4.3 Apparatus :

All stimuli were generated digitally by additive synthesis on a PDP 11/73 clone computer operating at a sampling frequency of 20 kHz. A program entitled PSYACX (Lai et al., 1987) allowed components to be specified in terms of their frequency, phase, amplitude, duration and envelope characteristics. After digital-to-analog conversion, the stimuli

were further band-pass filtered between 100 Hz and10 kHz via a brickwall filter (Wavetek-Rockland 751A) . The response of the filter rolled off by115 dB/octave outside the passband. The stimuli were then amplified to a comfortable listening level (about 70 dB SPL) using digital attenuators in conjunction with power amplifiers, and were presented binaurally over headphones (AKG 141) to the ears of a listener seated in a sound-absorptive listening room.

The accuracy of the synthesized spectra was verified via a Hanning-window analysis using a Bruel & Kjaer 2033 single-channel spectrum analyzer. A Tektronix dual-trace oscilloscope was used to view the waveforms and to verify the accuracy of the temporal layout of the stimuli in a sequence.

4.4 Procedure :

4.4.1 Subjects:

Seven listeners with ages ranging from 21 to 41 years were recruited as subjects via a job advertisement (3 male: GC, LC, JO; 4 female: CS, LJ, JK, WK). Their hearing was audiometrically verified to be within normal limits. Six of these listeners had already served as subjects in Experiment 1. Two of the subjects (GC and WK) were professional musicians with advanced musical training. Subject JO was

active in amateur music performance and subject LJ had some training in dance. Subjects CS and LC were avid listeners of music but did not have any formal musical instruction. The seventh listener (JK) was a research assistant with considerable listening and choral-singing experience.

4.4.2 Task:

While proportional ratio changes in the frequency of components of a complex preserve its harmonicity, linear changes lead to the sound becoming inharmonic. Such stimuli are known to be associated with shifts in overall pitch (de Boer, 1976; Schouten et al., 1962), with timbral percepts such as "roughness" (Plomp, 1970), and with the perception of multiple sources (Cohen, 1982; McAdams, 1984 a,b).

Given this knowledge *a priori*, a double-judgment task was set before the listeners. First, they were required to comment on the degree of fusion of the complex tones in the stimulus, and secondly, on any change perceived in the overall pitch or timbre of the sounds. This was accomplished by asking listeners to make two judgements for every stimulus. The instructions were phrased as follows:

"Upon hearing two repetitions of a pair of sounds, you will be required to make **two judgments** on each trial:

First: "Do the tones appear to "segregate" or "split" into more than one entity ? If so, which of the tones splits ? (1) first tone, (2) second

tone, (3) neither, (4) both ? Please press one of the keys labelled '1', '2', '3' or '4' on the computer keyboard to indicate your choice.

Second: "Is the second tone in the pair (1) same, (2) higher in pitch, (3) lower in pitch, (4) different in "something else", (5) different in "something else" and higher, (6) different in "something else" and lower in pitch than the first tone ? Please press one of the keys labelled '1', '2', '3', '4', '5' or '6' on the computer keyboard to indicate your choice." [As in experiment 1, "something else" is here considered to be synonymous with timbre, which was explained to listeners as being a qualitative feature of sounds distinct from their pitch, loudness or duration].

To help listeners remember the designated labels for the task, an illustration of the type shown in **figure 4.2** was provided during each session. An assortment of the stimuli was also presented to the subjects during initial briefing sessions to acoustically illustrate the range of percepts.

4.4.3 Stimulus presentation:

The same program (PSYACX) was used to run the experiment. Stimuli were assigned to blocks with number of replications specified for each stimulus. The stimuli in a block were chosen to have the same type of shift (i.e. either harmonic or inharmonic) for the different magnitudes

of frequency change. There were thus 4 such blocks for the 2 values of F0 for each type of change. Each block comprised 13 different stimuli having different amounts of frequency change (6 magnitudes X 2 directions (up/down) plus 1 unchanged, standard). The stimuli were replicated 4 times in a block in a randomized order, to yield a total of 52 trials per block.

After presentation of the two repetitions of the stimulus, the listener was given 3.5 seconds in which to make a response. A "response" comprised sequential pressing of 2 keys on a computer keyboard to indicate the choice of labels for the 2 judgements required.

Thus, the listener selected one of the four keys 1, 2, 3 or 4 for the first judgement pertaining to "splitting" of the tones, and one of the six keys 1, 2, 3, 4, 5 or 6 for the second judgement pertaining to the pitch and timbre of the tones. Upon elapse of the 3.5 second interval, a written message on the computer screen prompted the listener to make a response. Listeners were given short breaks between blocks.

A trial-by-trial record of the stimuli presented and corresponding responses was maintained by the computer. At the end of a block, the data could be retrieved in a matrix showing the set of stimuli presented in the block and the proportion of times different response keys were selected across replications of the stimuli.

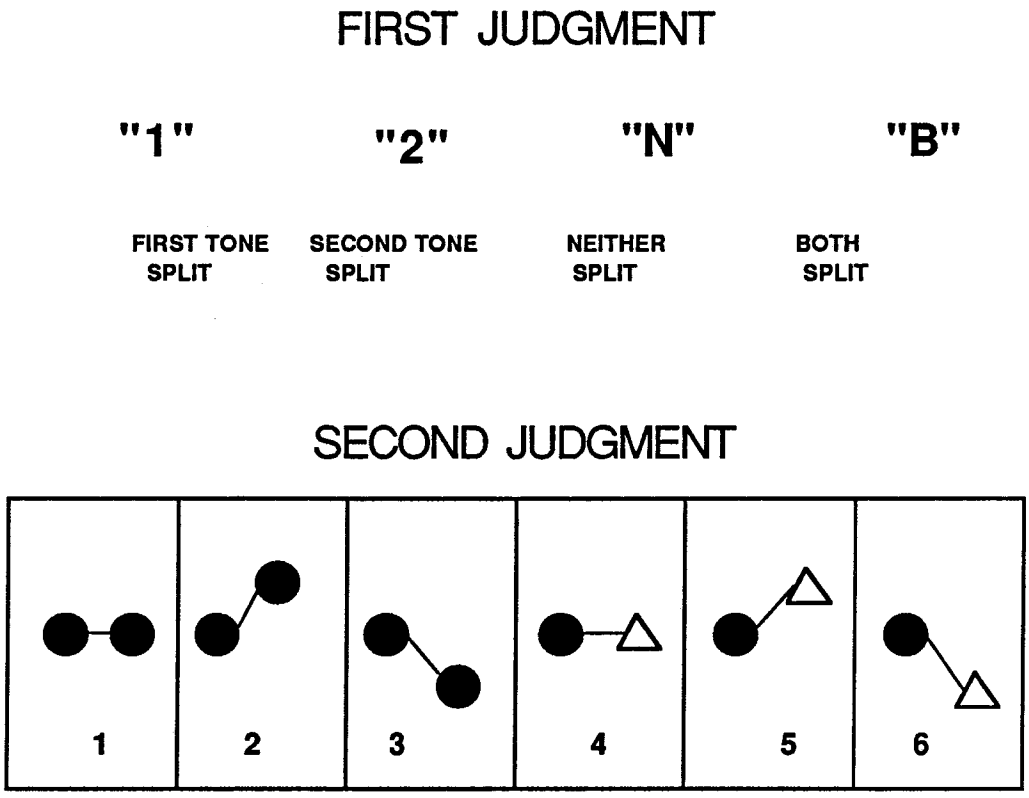


Figure 4.2 Double-task assigned in experiment 2. The first judgment pertained to perceived splitting or fusion of the tones. The four response keys corresponded to perceived splitting of 1) the first tone (label '1'), 2) the second tone (label '2'), 3) neither tone ((label 'N'), or 4) both tones(label 'B') The second judgment pertained to perceived differences in the pitch and timbre of the tones as illustrated.

4.5 Results and Discussion :

The numbers of times particular labels were selected by listeners to indicate their judgements in the 2-layer task constitute the "raw" data. These data have been processed in two different ways:

i) In order to show the general trend and **magnitude of use** of different labels, proportions of label use have been averaged for each response category across the 7 listeners. Figures 4.3 and 4.4 show the averaged data for the first and second judgements for reference F0=200 Hz for harmonic and inharmonic stimuli respectively. Figures 4.5 and 4.6 show similar data for the 400-Hz standard.

ii) In order to show **listener variability** and individual differences in response strategy, the dominant response labels for different magnitudes of frequency change for the different conditions are displayed in figures 4.7, 4.8, 4.9 and 4.10 in terms of the numbers of listeners who selected them.

A key to symbols used in the graphs appears in each figure. The labelling data obtained for the harmonic and inharmonic stimuli with ratio and linear changes in frequencies of components, respectively, are discussed separately below.

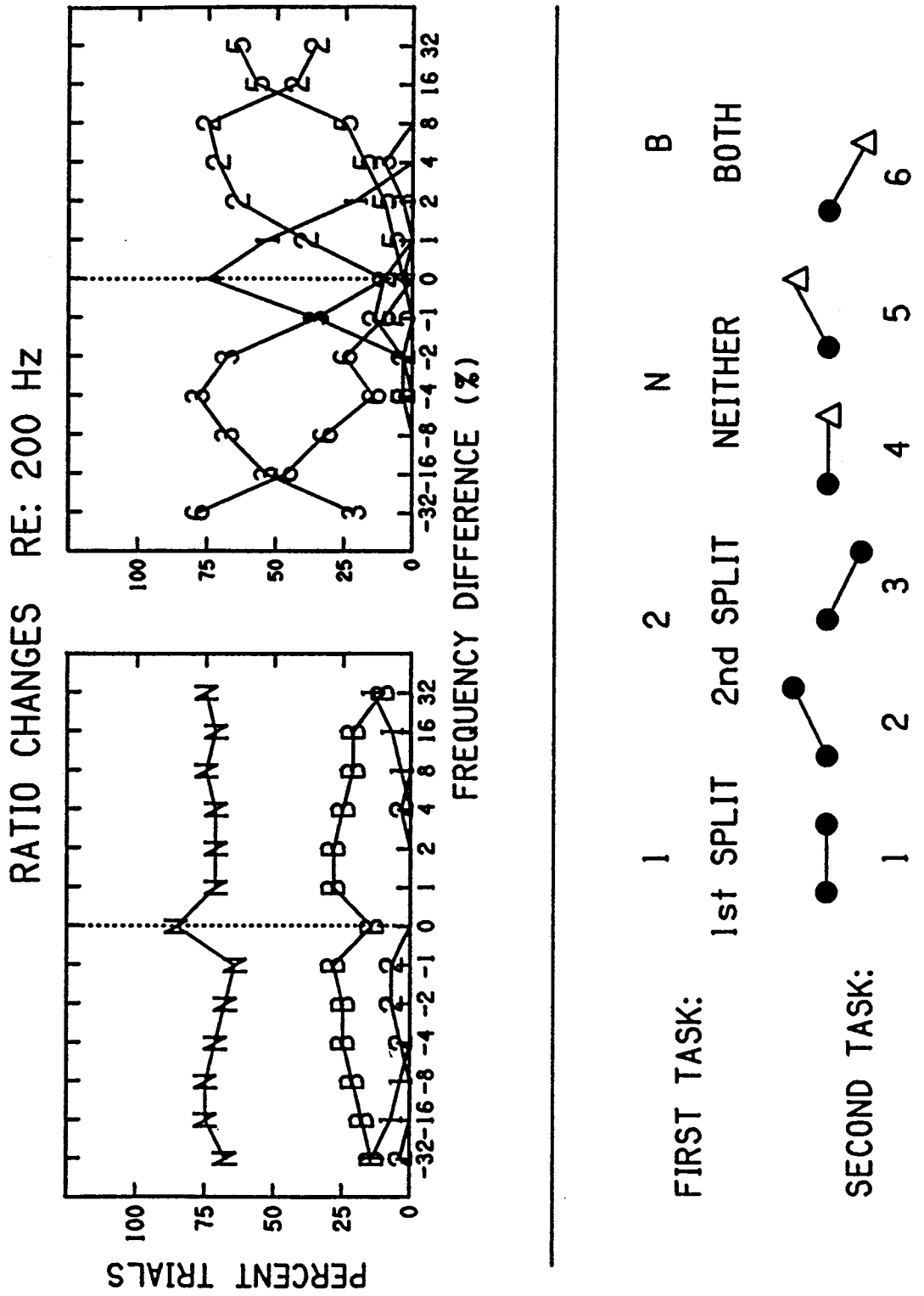
4.5.1 Magnitude-of-use of different response labels

4.5.1.1 Ratio changes with standard F0=200 Hz

The labels selected for stimuli with ratio changes of different magnitude averaged across 7 listeners, are shown in **figure 4.3.** In the left panel, the abscissa shows different magnitudes of ratio change (-Δf, 0, +Δf %), while the ordinate gives the proportion of use of the 4 labels available for the first judgment pertaining to fusion of the sounds. The right panel shows similar usage of labels for the second judgment pertaining to perceived pitch and timbre relations between the same sounds as were judged for the fusion data of the left panel.

For ratio changes that preserve harmonicity, one would expect the tones to remain fused and the label 'N' (key '3') to be selected 100% of the time. As can be seen in figure 4.3, the tones *were* indeed predominantly perceived as being fused, with label 'N' selected around 75% of the time. However, the use of other labels (notably 'B') on about 25% of trials, was also observed. This use of the 'B' response category is attributable to the performance of 2 of the 7 listeners (GC and LC), who appeared to hear the stimuli "analytically" as chord-like stacks of components, rather than as fused entities as perceived by most listeners. The variability of responses across listeners is discussed in greater detail in following sections.

Figure 4.3 (next page) Data for stimuli with ratio changes relative to F0=200 Hz. Magnitude of change is shown along the abscissa (in Δf %), and magnitude of use of different labels along the ordinate (in terms of the proportion of total trials). Response labels for the first judgment are shown on the left and for the second judgment on the right. The key under the graph illustrates the meaning of data symbols.


RATIO CHANGES RE: 200 Hz
PERCENT TRIALS
FREQUENCY DIFFERENCE (%)
FIRST TASK: 1 2 N B
1st SPLIT 2nd SPLIT NEITHER BOTH
SECOND TASK: 1 2 3 4 5 6

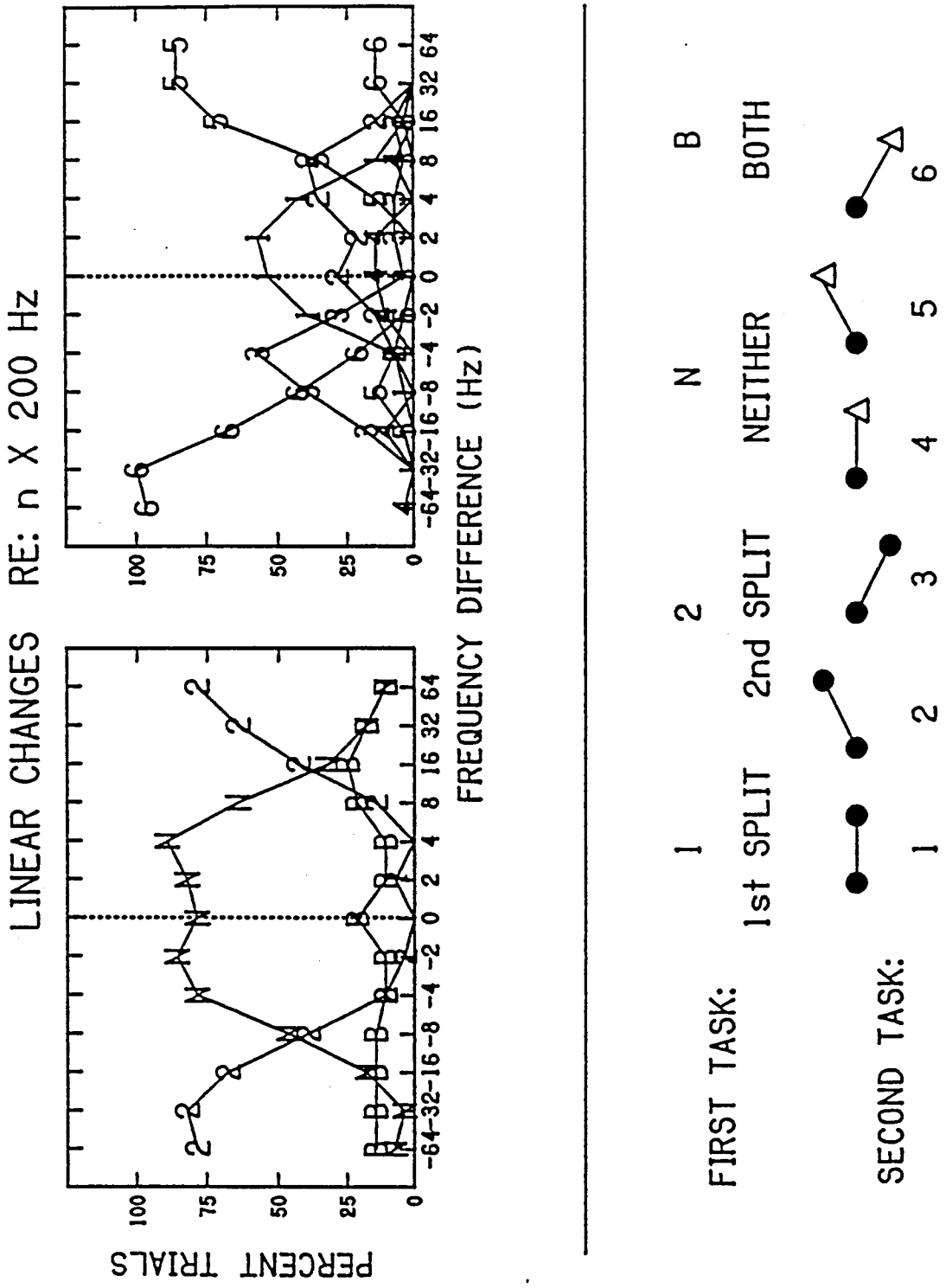
For the second judgment, the dominant use of labels '2' and '3' (shown in the right panel of figure 4.3) indicate that these ratio changes primarily evoked a change in pitch without an accompanying change in timbre. At the extreme ends of the Δf range however, the labels '6' and '5' were also used, implying that timbre changed in addition to pitch. Changes of this magnitude ($\approx$32%) in these sounds with flat spectral envelopes, atypical of most natural sounds, apparently led to differences in perceived timbre.

4.5.1.2 Linear changes with standard F0=200 Hz

For linear changes shown along the abscissae of the two graphs in **figure 4.4**, the first judgment data in the left panel show fusion for values of $\Delta f \leq 8$ Hz. However, the dominant response label changed from 'N' to '2' for greater magnitudes of Δf, implying perceptual segregation or "splitting" of the second tone. The label 'B' implying splitting of both tones was also used for these inharmonic stimuli on some trials (less than 25%), again contributed by 2 of the 7 listeners (GC and LC).

For the second judgment data shown in the right panel, a change in overall pitch was reported for linear changes of the order of $\pm$ 4 to 8 Hz in the frequency of spectral components (manifested by use of labels '2' and '3'). Larger deviations were construed as changes in **both pitch and something else** (labels '5' and '6').

Figure 4.4 (next page) Data for stimuli with linear changes given standard F0=200 Hz. Magnitude of change is shown along the abscissa (in Δf Hz), and magnitude of use of different labels along the ordinate (in terms of the proportion of total trials). Response labels for the first judgment are shown on the left and for the second judgment on the right. The key under the graph illustrates the meaning of data symbols.


LINEAR CHANGES RE: n X 200 Hz
PERCENT TRIALS
FREQUENCY DIFFERENCE (Hz)
100
75
50
25
0
-64 -32 -16 -8 -4 -2 0 2 4 8 16 32 64
1 2 N B
FIRST TASK: 1st SPLIT 2nd SPLIT NEITHER BOTH
SECOND TASK: 1 2 3 4 5 6

4.5.1.3 Proportion data for stimuli with standard F0=400 Hz.

Magnitude of use of response labels for harmonic and inharmonic stimuli for the 400-Hz condition are displayed in **figures 4.5 and 4.6** respectively. The results are similar to those obtained for the 200-Hz condition. The harmonic stimuli were perceived to be fused on ≈75% of the trials, with changes in F0 construed as changes in pitch (figure 4.5). Some use of the 'B' label was observed for the first judgment and changes of magnitude ≈32% were sometimes reported as changing the pitch as well as "something else" in the second judgment.

The inharmonic stimuli were perceived as being "split" given linear deviations in frequency $\geq \pm 16$ Hz (figure 4.6). However, analytic listening appeared to be enhanced for these stimuli, as manifested by a higher proportion of 'B' responses than was seen for the 200-Hz inharmonic stimuli (> 25%).

For the second judgment, pitch changes without accompanying changes in "something else" were reported for small deviations from harmonicity ($\Delta f \approx \pm 8$ Hz). Larger deviations were construed as changing both the pitch and "something else".

Figure 4.5 (next page) Data for stimuli with ratio changes relative to F0=400 Hz. Magnitude of change is shown along the abscissa (in Δf %), and magnitude of use of different labels along the ordinate (in terms of the proportion of total trials). Response labels for the first judgment are shown on the left and for the second judgment on the right. The key under the graph illustrates the meaning of data symbols.

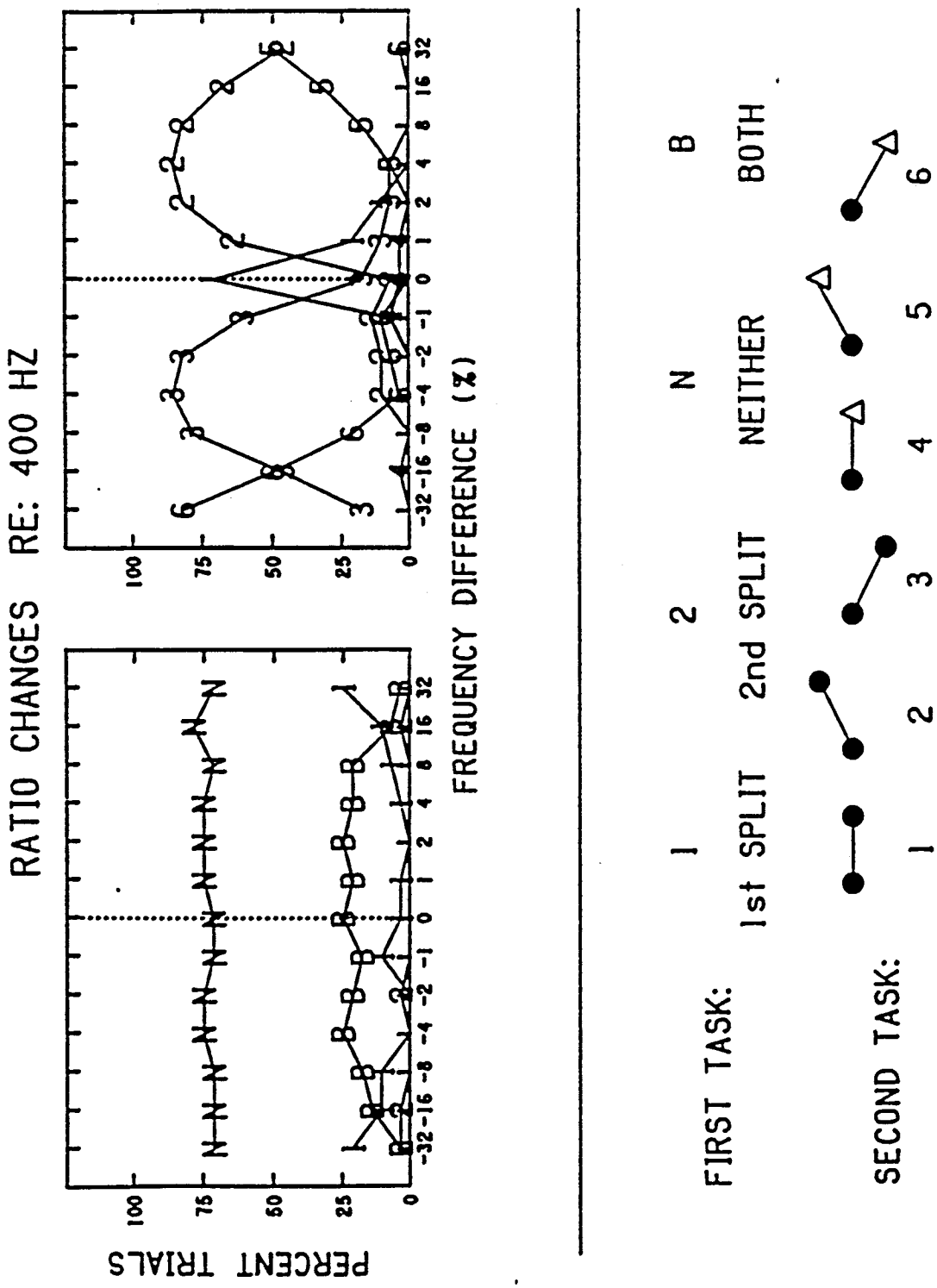
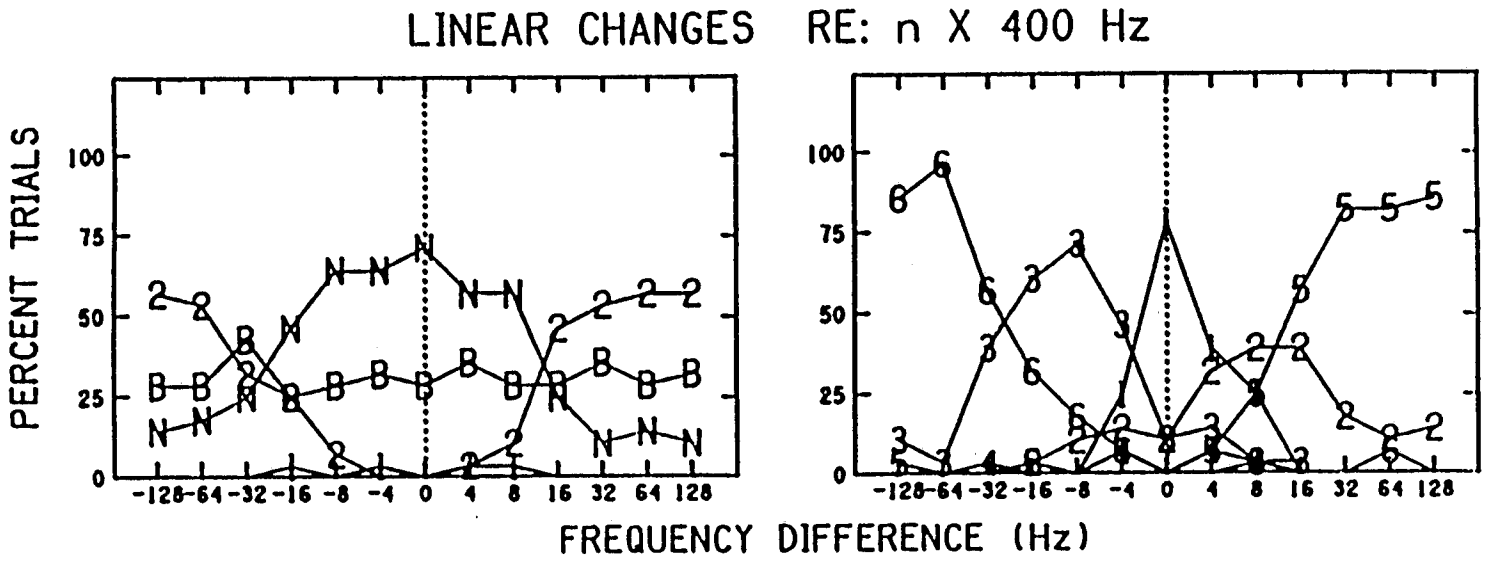
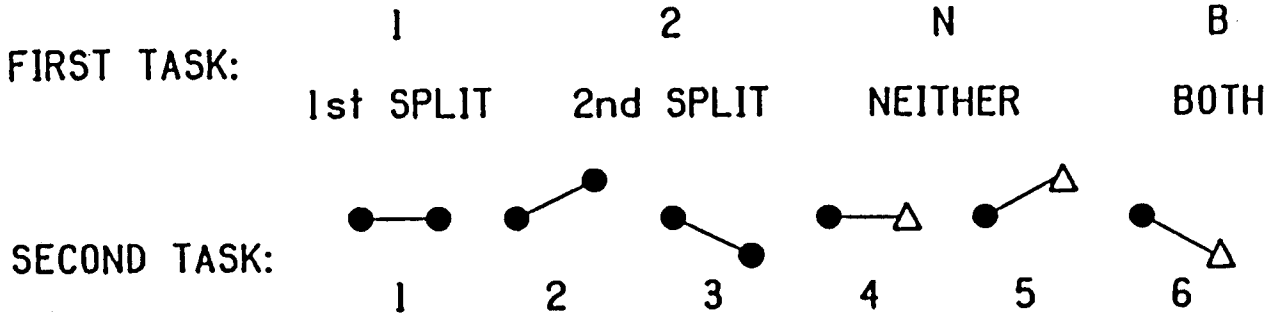

RATIO CHANGES RE: 400 HZ
PERCENT TRIALS
FREQUENCY DIFFERENCE (%)
1 2 N B
FIRST TASK: 1st SPLIT 2nd SPLIT NEITHER BOTH
SECOND TASK: 1 2 3 4 5 6

Figure 4.6 (Next page) Data for stimuli with linear changes given standard F0=400 Hz. Magnitude of change is shown along the abscissa (in Δf Hz), and magnitude of use of different labels along the ordinate (in terms of the proportion of total trials). Response labels for the first judgment are shown on the left and for the second judgment on the right. The "key" illustrates the percepts implied by the response labels. The key under the graph illustrates the meaning of data symbols.


LINEAR CHANGES RE: n X 400 Hz
PERCENT TRIALS
100
75
50
25
0
-128 -64 -32 -16 -8 -4 0 4 8 16 32 64 128
FREQUENCY DIFFERENCE (Hz)


1
2
N
B
FIRST TASK:
1st SPLIT
2nd SPLIT
NEITHER
BOTH
SECOND TASK:
1
2
3
4
5
6

4.5.2 Listener variability

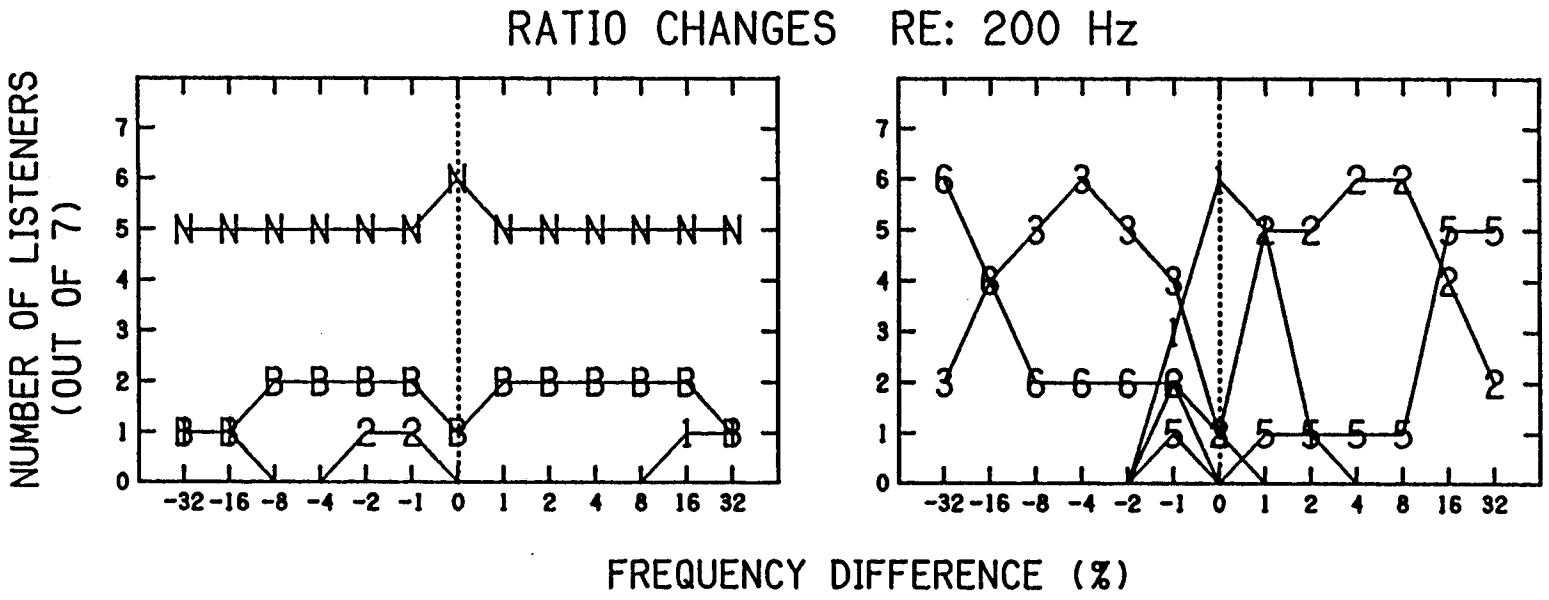
Given differences in response across listeners, particularly in regard to the use of label 'B' for the first judgment, the labelling data are re-plotted in figures 4.7 through 4.10 for the different conditions, in terms of **numbers of listeners** and their preferred response labels.

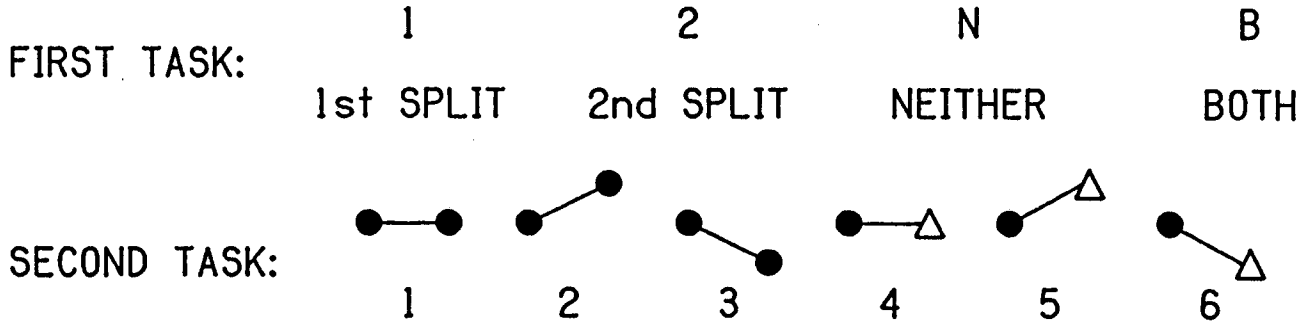
4.5.2.1 Ratio changes

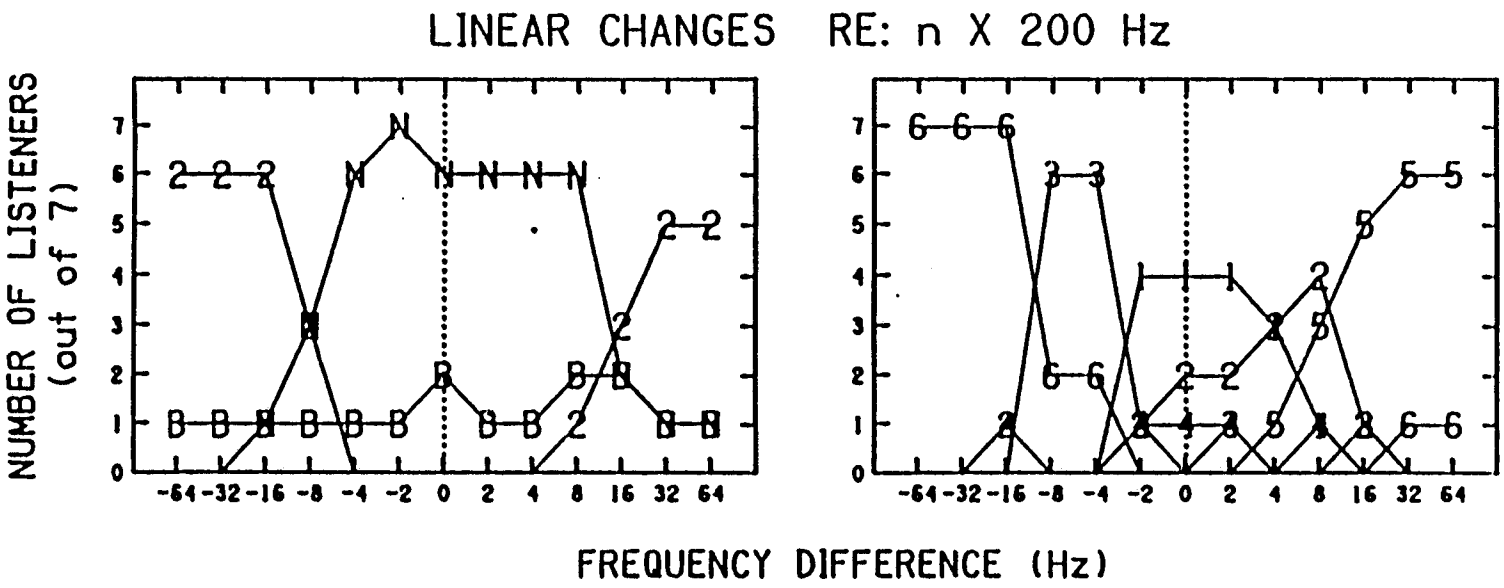
The distribution of response labels across numbers of listeners can be seen in **figure 4.7** for the two judgments for ratio changes with standard F0= 200 Hz. Again, magnitude of change is given along the abscissa. The data points represent the response options available, with the number of listeners (out of 7) who selected them indicated along the ordinate. Equivocation between choice of labels is manifested by the total number of listeners (added across labels) exceeding 7 for a given magnitude of change.

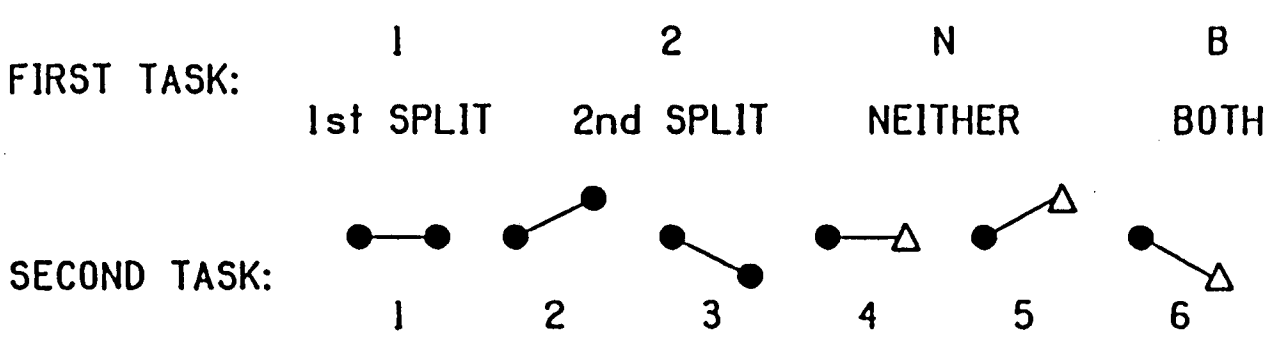
For the first judgment, the left panel of figure 4.7 shows that harmonic stimuli were consistently perceived as being fused by at least 5 out of the 7 listeners. The other two listeners (GC and LC) appear to have operated in an enhanced "analytic" mode, reporting both sounds, and sometimes even the first (unchanged) sound alone as being "split" !

Figures 4.7 and 4.8 (Next two pages) Distribution of dominant response labels across listeners for stimuli with ratio changes relative to F0=200 Hz and 400 Hz, respectively. Magnitude of change is shown along the abscissa (in Δf %), and the number of listeners using different labels is shown along the ordinate. Response labels for the first judgment are shown on the left and for the second judgment on the right. The key under the graph illustrates the meaning of data symbols.


RATIO CHANGES RE: 200 Hz
NUMBER OF LISTENERS (OUT OF 7)
FREQUENCY DIFFERENCE (%)
-32 -16 -8 -4 -2 -1 0 1 2 4 8 16 32


FIRST TASK:
1
1st SPLIT
2
2nd SPLIT
N
NEITHER
B
BOTH
SECOND TASK:
1
2
3
4
5
6


LINEAR CHANGES RE: n X 200 Hz
NUMBER OF LISTENERS (out of 7)
FREQUENCY DIFFERENCE (Hz)
-64 -32 -16 -8 -4 -2 0 2 4 8 16 32 64


FIRST TASK:
1
2
N
B
1st SPLIT
2nd SPLIT
NEITHER
BOTH
SECOND TASK:
1
2
3
4
5
6

Listener GC is a professional musician with extensive experience in ear training and "musicianship". His ability to listen analytically thus does not seem unusual. However, LC is not formally trained as a musician, yet he often used the 'B' label as well. The other musician (WK), on the other hand, did not report fission of the first tone and followed the mean trend of response exhibited by the other subjects.

For the second judgment (right panel of figure 4.7), listeners were in good agreement with each other, with 5 to 6 listeners consistently reporting pitch changes with change in F0. For the largest values of change (32%), 2 listeners (GC and LJ) disagreed with the majority, who reported changes in "something else" in addition to changes in pitch.

A very similar distribution of responses was observed for the harmonic stimuli with reference F0=400 Hz (**figure 4.8**).

4.5.2.2 Linear changes

Distribution of response labels across listeners for the two judgments for inharmonic stimuli with standard F0=200 Hz is shown in **figure 4.9**. There was good agreement between listeners for the first judgment (shown in the left panel). Almost all listeners (6 to 7 out of 7) used the label 'N' to indicate lack of splitting, (or fusion) of sounds over the range of changes from $\Delta f=0$ to ± 8 Hz. For larger deviations, 5 to 6 listeners perceived the second (inharmonic) sound as being split. Listener GC (and sometimes LC) continued to report both sounds as being

split.

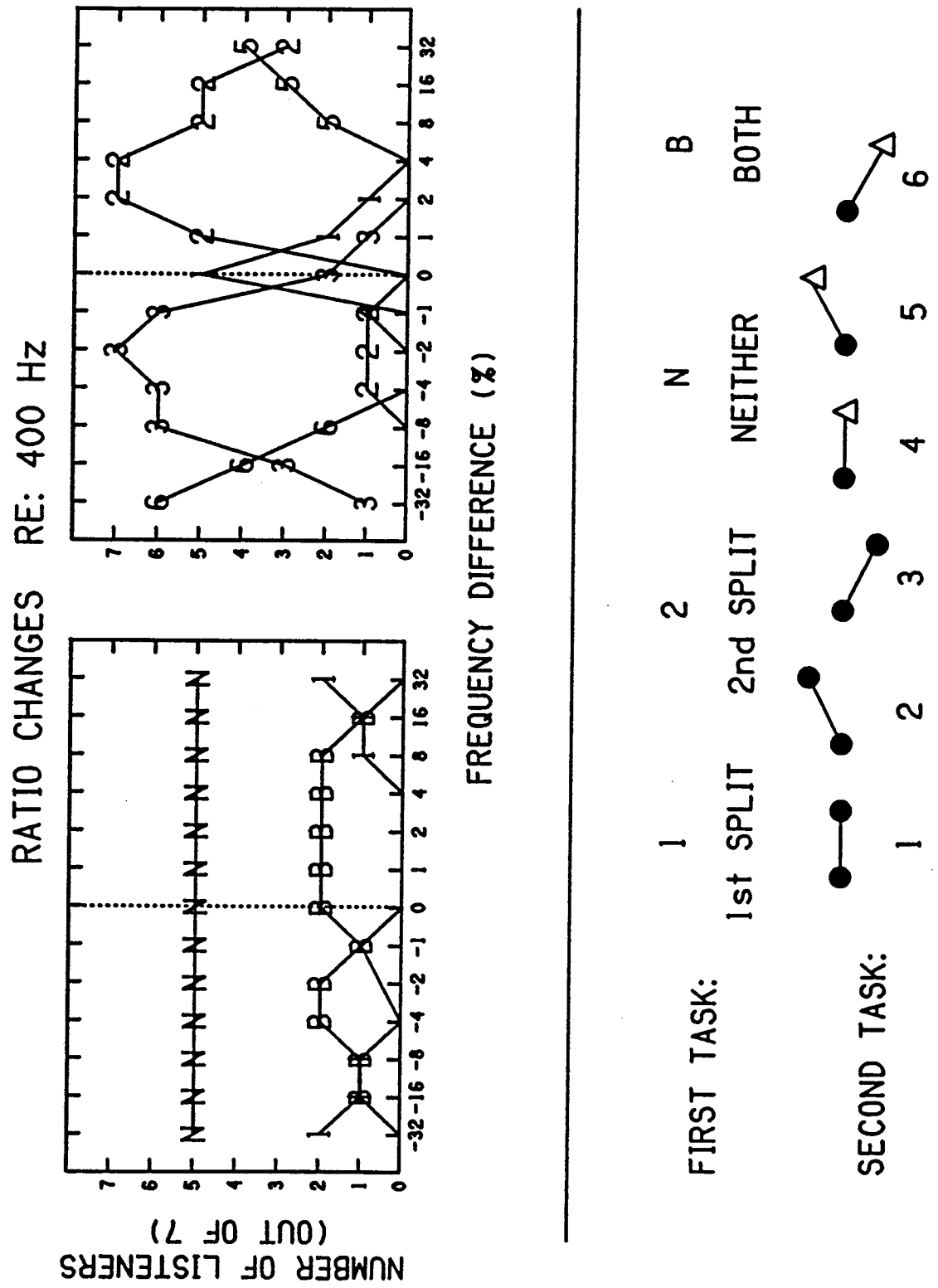
For the second judgment, listeners were in good agreement, reporting changes in pitch along with changes in "something else" for $\Delta f \geq 8$ Hz (5 to 7 listeners out of 7). For smaller deviations, other response labels were also used, though the majority of listeners agreed in reporting no change for small shifts (0 to 4 Hz) and changes in pitch only, for $\Delta f \approx 4$ to 8 Hz.

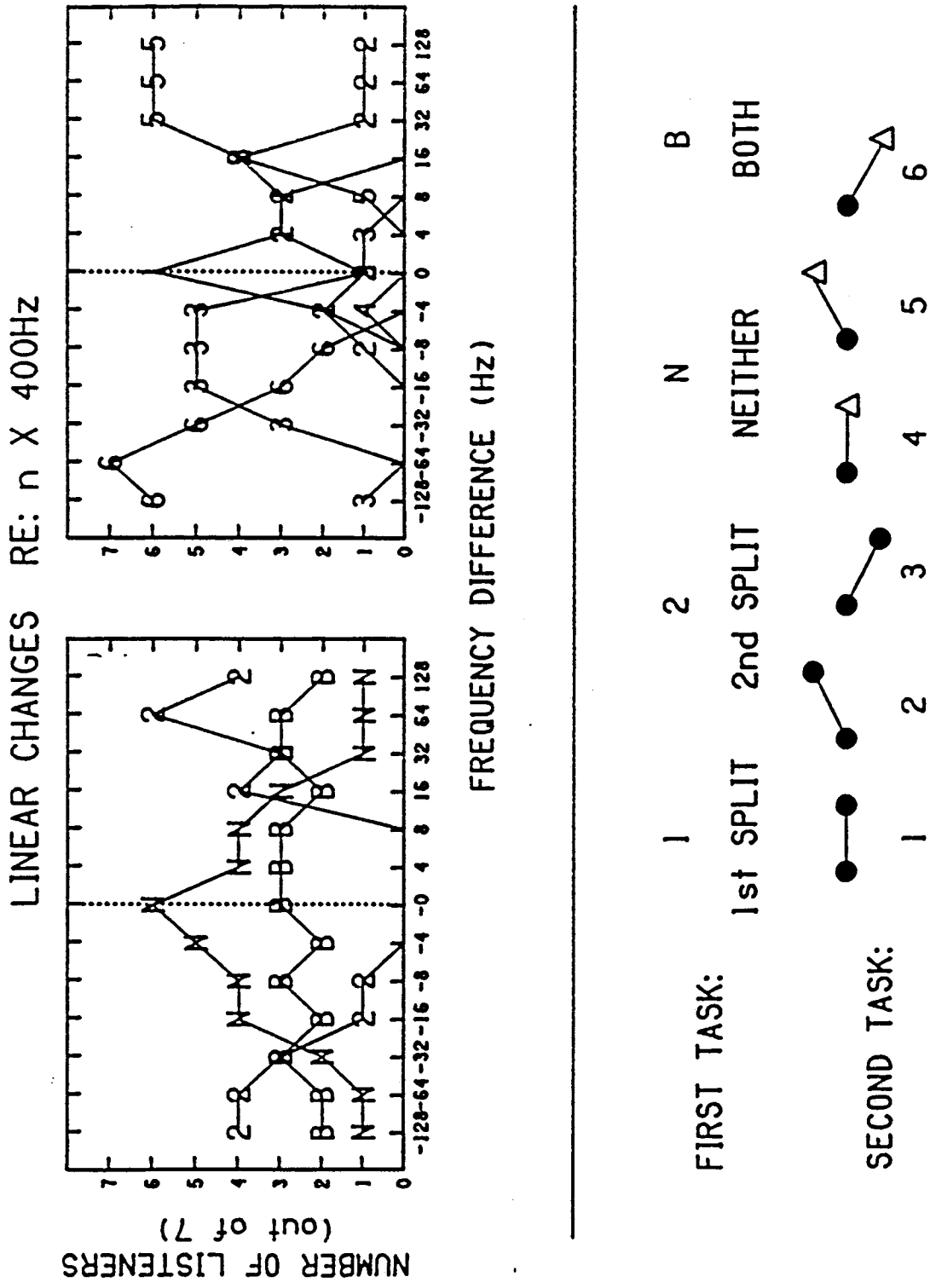
Distribution of response labels across listeners for the two judgments for inharmonic stimuli with standard F0=400 Hz is shown in **figure 4.10**. The responses for the first judgment are seen to be more variable for this reference frequency, than they were for the analogous condition at 200 Hz (figure 4.9).

A larger number of listeners reported splitting of both tones (3 instead of 2; GC, LC and JO), and one listener (JK) continued to report fusion even for large changes in frequency. The major response trends noted from the proportion data are still visible however, with fusion being dominant for $\Delta f \leq 16$ Hz and splitting reported for greater deviations from harmonicity.

For the second judgment, this variability across listeners was not observed. On the contrary, 5 to 7 (out of 7) listeners agreed in reporting pitch changes for $\Delta f \leq 16$ Hz, and changes in both pitch and "something else" for larger values.

Figures 4.9 and 4.10 (Next two pages) Distribution of dominant response labels across listeners for stimuli with ratio changes relative to F0= 200 Hz and 400 Hz, respectively. Magnitude of change is shown along the abscissa (in Δf %), and the number of listeners using different labels is shown along the ordinate. Response labels for the first judgment are shown on the left and for the second judgment on the right. The key under the graph illustrates the meaning of data symbols.


RATIO CHANGES RE: 400 Hz
NUMBER OF LISTENERS (OUT OF 7)
FREQUENCY DIFFERENCE (%)
-32 -16 -8 -4 -2 -1 0 1 2 4 8 16 32
0 1 2 3 4 5 6 7
N N N N N N N N N N N
B B B B B B B B B B B
FIRST TASK: 1 2 N B
1st SPLIT 2nd SPLIT NEITHER BOTH
SECOND TASK: 1 2 3 4 5 6


LINEAR CHANGES RE: n X 400Hz
NUMBER OF LISTENERS (out of 7)
FREQUENCY DIFFERENCE (Hz)
-128 -64 -32 -16 -8 -4 -0 4 8 16 32 64 128
0 1 2 3 4 5 6 7
FIRST TASK: 1st SPLIT 2nd SPLIT
1 2 N B
NEITHER BOTH
SECOND TASK: 1 2 3 4 5 6

4.6 General Discussion :

Consideration of both absolute and relative aspects of the spectrum in derivation of pitch was indicated by the results of the first experiment. This was further verified in the present experiment. The dependence of other percepts such as timbre and perceived fusion on relations between frequencies of components was also observed.

In the present experiment, the spectral spacing of components was varied in ways that preserved or disrupted the harmonicity of a complex sound. Proportional changes in the frequencies of components were generally perceived as changes in pitch of a fused complex tone. Deviations from harmonicity led to a compound of possible perceptual changes, including changes in **pitch**, changes in **timbre**, and loss of **fusion**.

4.6.1 Changes in pitch of inharmonic sounds

Inharmonic stimuli are typically associated with ambiguous pitches (Cohen, 1980; Terhardt, 1974). However, changes in pitch were reported for almost all values of Δf used in the present experiment. The perception of pitch for these inharmonic stimuli may be a consequence of their special nature, in that the components were equally spaced in terms of absolute frequency differences.

Changes in pitch for inharmonic signals of this type have been studied extensively in the past (Schouten et al., 1962; de Boer, 1956/1976). Both spectral and temporal theories have been proposed to account for the shift in pitch often observed for such stimuli where spectral spacing is preserved, but not harmonicity:

Schouten et al. (1962) suggested that a "pitch extractor" operates in the time domain, measuring time intervals between both major and minor peaks in the "fine structure" of the waveform. Pitch is then derived from the inverse of the "pseudo-period" as was originally proposed by de Boer (1956/1976).

In light of data on dichotic pitch (Houtsma and Goldstein, 1972), temporal theories of this type based on perception of relations in the "beat" pattern produced by interfering components, have now largely been abandoned in favor of spectral pattern-matching theories (Goldstein, 1973; Duifhuis, Willems and Sluyter, 1982; Grandori, 1984). The inclusion of combination tones and relative dominance of lower harmonics in the pitch estimation process have been considered as contributory to the pitch shifts observed for inharmonic complexes with constant spacing.

The stimuli used in our experiments were presented diotically (to both ears) at fairly intense levels (≈70 dB SPL). Both temporal interference and combination tone phenomena may have been operative under these conditions. Such consequences were anticipated in the

design of the stimuli. Since the aim of the experiment was not to study "mechanisms" of pitch coding, where such factors may be confounding, they were not deliberately avoided, as is sometimes the case. Rather, the stimulus level and manner of presentation was chosen to be representaive of routine listening situations.

The aim of the experiment was not to obtain pitch matches or estimates of pitch shifts, but rather, to note **if** pitch shifts occurred at all, and to determine the range of frequency changes over which they were perceived. For small deviations from harmonicity (≤4% of spectral spacing), pitch changes were reported without any accompanying changes in "something else". Larger deviations, however, induced reports of change in both **pitch, and "something else"**. Taking the term "something else" to be synonymous with timbre, the consequences of deviations from inharmonicity on the perception of timbre are discussed in the next section.

4.6.2 Changes in timbre

The early experiments on pitch shifts did not indicate that changes in timbre may have accompanied the stimulus changes. In the present experiment however, the use of labels '4', '5' and '6' for the second judgment and verbal reports by subjects describing changes in quality of the sounds indicated that timbre changes **did** occur.

Both spectral and temporal factors may be responsible for the perceived changes in timbre. The changes in spectral spacing led to components being shifted in frequency and to changes in spectral bandwidth. Since timbre is known to be correlated with the absolute position of the spectrum, such changes may well be perceived as changes in timbre.

The displacement of components also leads to altered spectral spacing. Changes that lead to reduced inter-component spacing may lead to interference within critical bands that is manifested as a timbre change. The10 components of the inharmonic stimuli spanned the range 136 to 4128 Hz for reference F0=200 Hz, and 272 to 4128 Hz for reference F0=400 Hz. Critical bandwidth estimates for this range vary from 100 Hz to 500 Hz (Moore, 1982). The reduced spacing for negative values of Δf (-8 to -32% of the spacing) could lead to the higher components of the complex sound interfering within critical bands.

Inharmonic stimuli of the type used here are also associated with dynamic **changes in the waveform** caused by drifting phase relations between "mistuned" components. The multiple periodicities indicated by components that are not harmonically related lead to a loss of consistent phase structure both in the physical stimulus, and in the response patterns observed at higher levels of auditory processing (Plomp, 1970). The modulations of the waveform caused by ongoing alterations in the phase pattern of components could well be construed as changes in

timbre.

Figure 4.11 shows waveforms for one of the sequences used in the present experiment (sharpness of the peaks is an artifact of the digital scaling program used). The two sounds of a stimulus pair are shown side by side for comparison (not to scale in time). The waveform on the left is for the "standard" harmonic complex tone with F0=200 Hz. The waveform on the right is for an inharmonic complex with all components shifted linearly in frequency by 32 Hz relative to the components of the standard.

The harmonic sound exhibits regular periodicity, while modulation of the inharmonic sound is clearly visible. The type of phase effect described by Plomp (1970) in reference to timbre changes, may thus indeed have been operative in the present experiment.

Phase effects in audition have been studied by many other investigators as well (Buunen et al., 1974; Fischler, 1967; Goldstein, 1967; Mathes and Miller, 1947; Zwicker, 1952). The "discrimination" paradigms under which these were usually studied did not address the issue of timbre perception directly. However, the perceptual adjectives often used to describe the stimuli encompass timbral attributes such as "roughness", "buzzing", "dissonance", "flutter" and other changes in "quality".

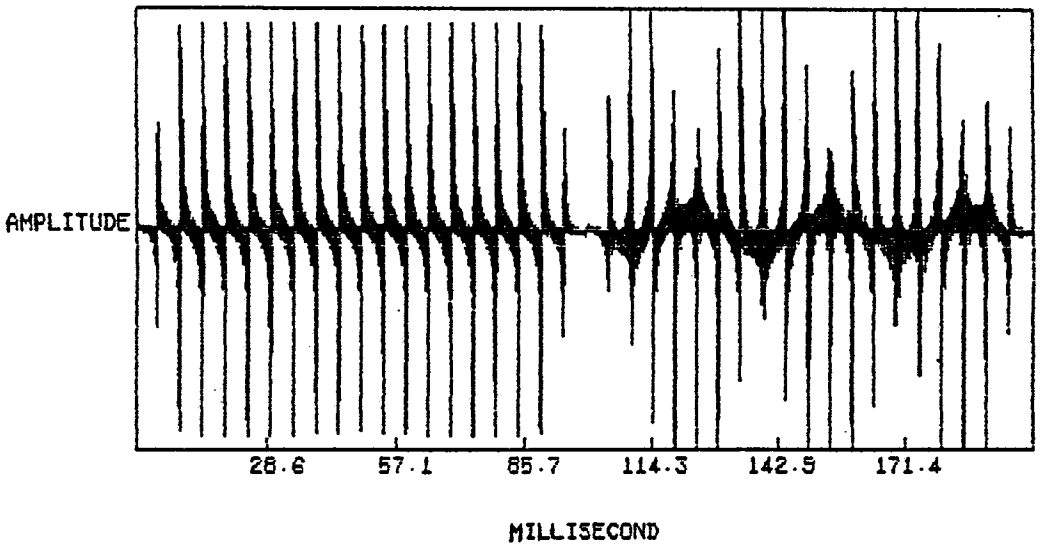


Figure 4.11 Waveforms for one of the sequences used in the present experiment (sharpness of the peaks is an artifact of the digital scaling program used). The two sounds of a stimulus pair are shown side by side for comparison (not to scale in time). The waveform on the left is for the "standard" harmonic complex tone with F0=200 Hz. The waveform on the right is for an inharmonic complex with all components shifted linearly in frequency by 32 Hz relative to the components of the standard. The harmonic sound exhibits regular periodicity, while modulation of the inharmonic sound is clearly visible.

The present experiment was designed from a "spectral" point of view, in that perceptual correlates of frequency changes of different magnitude were investigated under different contexts. It seems however, that the temporal consequences of these spectral changes contributed to percepts such as timbre.

4.6.3 Loss of fusion

The "classic" experiments on pitch shifts also neglected to mention that the inharmonic signals being used could lose their unified image and be perceived as a diffuse collection of sources (Cohen,1980, McAdams, 1984 a,b). The importance of harmonicity as a cue for fusion was mentioned in chapter 2. The perceived splitting of sounds in the present experiment seems to be a direct consequence of defiance of harmonicity. Had other compensating cues been provided, such as correlated modulation in frequency and/or amplitude (McAdams, 1984 a,b; Bregman et al., 1990), or the imposition of an exponential temporal envelope (Cohen, 1980), perhaps even these inharmonic stimuli would have remained fused.

The lack of fusion associated with the type of stimuli used here, may also account for the observed elevation of signal-detection thresholds when tones are embedded in some inharmonic maskers (Kohlrausch and Jacobi, 1989). **Figure 4.12** illustrates the stimulus

used by these authors. The first 210 msec of the masker was a periodic complex tone with the 10th to 30th harmonics of 50 Hz added in zero phase. The components were shifted by ≈4 Hz during the last 190 msec of the masker. The masked threshold for a10 msec signal embedded in the last 15 msec of the periodic part of the masker was observed to be 20 dB greater than that for a harmonic tonal masker. The authors attributed this rise to the deviation of the masker's temporal envelope from the regularity observed for harmonic complexes. They also reported that extensive training with the stimulus led to better performance and thus touted the "great influence of central stages of the hearing system even in simple detection tasks" (p.259).

It is suggested that the elevated thresholds obtained for the inharmonic stimulus were a consequence of increased difficulty in discriminating the masker with the presence of an additional source (the "signal") from the masker alone, since the latter could *a priori* be perceived as having multiple sources, by virtue of the induced inharmonicity.

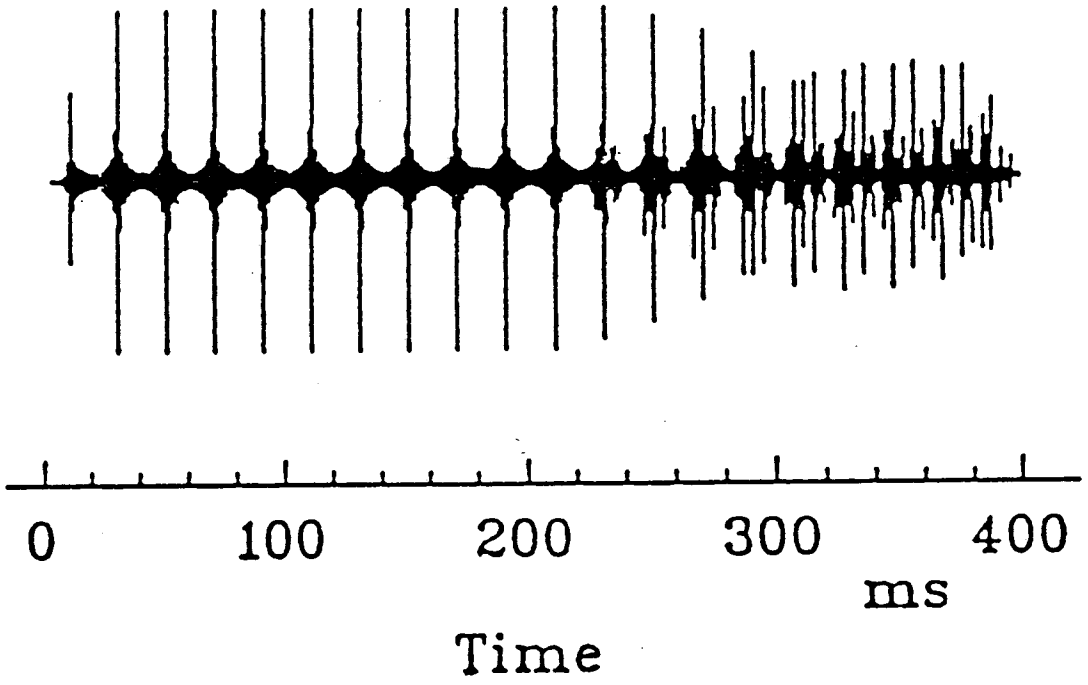


Figure 4.12 Waveform of an inharmonic masker used by Kohlrausch and Jacobi (1989). The first 210 msec of the masker was a periodic complex tone with the 10th to 30th harmonics of 50 Hz added in zero phase. The components were shifted by ≈4 Hz during the last 190 msec of the masker. The masked threshold for a10 msec signal embedded in the last 15 msec of the periodic part of the masker was observed to be 20 dB greater than that for a harmonic tonal masker. The authors attributed this rise to the deviation of the masker's temporal envelope from the regularity observed for harmonic complexes.

4.6.4 Comparison of results with McAdams (1984 b)

McAdams (1984 b) conducted a number of experiments on **spectral fusion and parsing** that employed stimuli very similar to those used in the present experiment. The frequencies of spectral components in his stimuli, however, were dynamically modulated, using 2 types of FM. One type preserved the relative frequency relations between components. All harmonics were modulated to an extent proportional to the harmonic frequencies. The harmonicity of the complex was thus preserved. Another type of FM preserved the spectral spacing between components. In this case, all components were modulated to the same absolute extent, and the complex thus became inharmonic.

For **ratio-preserving FM**, the stimuli retained their perceptually **fused** character and were heard for the most part as timbrally rich tones with varying pitch. For FM that **preserved spacing** between components but led to inharmonicity, the unitary image of the stimulus was lost and **multiple sources** perceived.

McAdams' listeners were presented a sequence of two sounds in the format of a "2I2AFC" task and were required to identify the interval that appeared to contain the greater number of sources. The "source multiplicity threshold" at which the inharmonic stimuli were reported as having more sources than the harmonic, occurred at values of Δf between

7 and 14 "cents" (≈0.4 to 0.8% change).

The changes employed in the present experiment ranged from 1% to 32% of the spectral spacing. These should all have been supra-threshold cues for segregation, yet our listeners reported fusion up until values of Δf=8%.

This difference in our results and those of McAdams could well be a consequence of the **different tasks** assigned. Our listeners had more options to describe their percepts, including the reporting of an **absence** of fission. McAdams listeners were obliged to choose the stimulus interval that contained "more sources". Stimuli that were different from the standard in ways other than "multiplicity of sources" could perhaps have been reported as segregated, given the lack of choice.

McAdams acknowledged that other perceptual changes were active and may have influenced judgments. His listeners were instructed to ignore the timbre differences that were sometimes discernible, but it was accepted that in some cases a **pitch-like criterion related to the width of modulation** could be used in making judgments.

McAdams' experiments focussed on **fusion and parsing** of complex sounds. However, the percepts he reports as being associated with component frequency changes are comparable to those observed here, namely: **changes in perceived fusion, pitch and timbre.** The present experiment validates the existence of all these types of percepts and shows the regions of frequency change where they come

into play.

4.7 Conclusions :

1). Harmonicity- preserving ratio changes are usually construed as changes in pitch. The changed sound retains its tonal quality and appears to be **fused** into a single entity.

2). For small deviations from harmonicity, of the order of $\Delta f=0$ to $\pm$ 8% of spectral spacing=200 or 400 Hz, the sounds are perceived as being fused, and changes in frequency are construed as changes in pitch.

3). Larger deviations from harmonicity lead to perceptual splitting of the inharmonic sound. The perceptual correlates of the frequency changes are reported as changes in pitch, accompanied by changes in timbre.

4.8 Epilogue

Since all components of the test sound in this experiment were changed simultaneously, the relative dominance of any particular component in evoking the associated perceptual change was not discernible. If the discrimination performance of a listener is in fact based on discriminability of the most-discriminable component as suggested by many researchers (e.g. Moore et al., 1984), then the relative

discriminability of different harmonics for equivalent amounts of frequency change needs to be investigated.

When the F0 of a harmonic complex tone is raised or lowered, all components comprising the tone get shifted in frequency as well. In terms of a ratio relative to the reference frequency, these changes are equivalent for all components. Thus for example, a 1% change in F0=200 Hz will result in a new F0=202 Hz. The frequency of the second harmonic will change from 400 Hz to 404 Hz (also equal to a 1% shift), the frequency of the third harmonic will change from 600 to 606 Hz (also equal to 1%) and so on. On a linear scale however, these shifts are different for the different harmonics (2 Hz for n=1, 4 Hz for n=2, 6 Hz for n=3, 8 Hz for n=4, and so on). Discrimination experiments employing ratio changes in harmonic complex tones may thus be providing multiple frequency-shift cues of differing magnitude.

The present experiment did not address this issue of relative contribution of different components to the percepts observed. To investigate the variation in percepts as a function of the component number being changed, a third experiment was conducted in which similarly dense spectra were used, but only a single component was changed in frequency. The task was again a double judgement pertaining to the three situations of interest, i.e. the perception of multiple sources, the perception of a pitch change and the perception of timbre changes. Details of the experiment are described in the following chapter.

CHAPTER 5

EXPERIMENT 3

Perception of complex tones with changes made in the frequency of single components

5.1 Introduction

Experiment 2 validated the existence of at least three types of percepts associated with changes in frequency that resulted in inharmonicity:

1). Shifts in perceived pitch

2). Changes in timbre

3). Loss of fusion

The cumulative experiments of Moore et al. (1984, 1985 a,b, 1986), McAdams (1984 a, b) and Hartmann (1988) with complex tones in which a single harmonic was "mistuned" also revealed similar perceptual correlates of frequency changes (see section 2.10 for details). However, in many of those experiments, the task focussed on a single percept (such as "hearing out" a harmonic), and did not allow other *covarying* perceptual changes to be reported.

The many experiments of Moore et al. demonstrated the importance of the task assigned to listeners when multiple cues are available for discrimination. Thresholds of different magnitude were

found for discrimination based on frequency change, depending on the task . Further, the function relating frequency DL (or a similar "threshold") to frequency region (in terms of harmonic number), varied for different tasks. Figure 2.11 in chapter 2 illustrates such differences in results, depending on the task assigned, given very similar stimuli and the same set of listeners.

While the frequency DL obtained via *pitch* change judgments and the frequency difference threshold for *"hearing out"* a component increased with rising harmonic number, thresholds for detection of *inharmonicity*, decreased. The perceptual changes used as criteria cueing discrimination (or identification) thus appeared to vary in salience, with harmonic number.

For the perception of pitch, it is known that different components of a complex sound carry different weights in contributing to the overall pitch percept (Goldstein, 1973; Plomp, 1967; Ritsma, 1967; Terhardt, 1974). Lower components, that are better "resolved" via cochlear filtering and usually have lower harmonic *numbers* than higher components, are considered to play a dominant role in the derivation of pitch from the spectrum.

For the perception of timbre, no general predictive scheme about relative contribution of components prevails. Timbre is a multidimensional percept dependent on both spectral and temporal factors. Changes in waveform, whether in periodicity or shape of the temporal envelope or in the fine structure, could well influence salience

of perceived changes of timbre. From a temporal point of view , one would expect **high components** to be greater contributors because of mutual interference within critical bands. Such interference is typically manifested in the structure of the temporal waveform and is perceived as "roughness" (see section 2.7.4). There is some indication that changes in "roughness" were operative in the task used by Moore et al. (1985 b) in which listeners were required to detect "inharmonicity" .

However, timbre is highly correlated with spectral "center of gravity" as well (von Bismarck, 1974; Grey, 1975; Wessel et al., 1987). A concept similar to that of the spectral "dominance region" for pitch may then exist for timbre as well. Different spectral regions may contribute differentially to changes in perceived timbre. If the dominance of these regions in *pitch* perception is based on the sensitivity of the ear to certain frequency ranges, then one would expect those same sensitive regions to contribute to timbre changes as well. From a spectral point of view, then, one would expect lower, better- resolved components to carry more weight in coveying timbre change.

A third type of change associated with inharmonic complexes is the lack of fusion. While experiment 2 verified that inharmonic changes lead to perceived segregation of a sound, the relative contributions of different components to this percept was not explored.

The extensive experiments of McAdams (1984 a,b) showed that deviations in the frequency of upper partials of a harmonic complex yielded *lower* "source multiplicity thresholds", corresponding to

magnitude of frequency change in different components, that led to the test sound being perceived as comprising many sources. Yet, another experiment by Moore et al. (1986) using this type of stimulus indicated somewhat *higher* thresholds for hearing "2 sources" as a function of harmonic number.

Again, it seems that the nature of the task assigned was critical to the results obtained despite the similarity of stimuli used. In McAdams' experiments, listeners were asked to judge which of two presentation intervals contained the sound with more apparent sources. In Moore et al.'s experiment, a single sound was presented and listeners were asked to report if they heard "a single sound with one pitch, or two sounds, -a complex tone and a component with a pure-tone quality not belonging to the complex". This criterion is more stringent than the one used by McAdams.

In McAdams' task, listeners were not explicitly asked to report *how many* sources were perceived. The criterion related to hearing more sources could therefore be loosely inferred by listeners, with percepts such as beating and roughness construed as being indicative of "multiple sources". Since such cues were surely available for the densely-packed high harmonics, low thresholds could be obtained based on ease of discriminability of the compared sounds, rather than on the basis of actually perceiving more than one source.

Moore et al.'s task on the other hand, set the listener up to ignore subtle effects such as roughness and respond positively only when there

was clear perceptual segregation of the sound into two definitely separate entities. Higher thresholds were thus obtained for the high harmonics than would be expected from the results of McAdams' experiment.

Hartmann (1988) also reported the existence of multiple percepts associated with mistuning of a harmonic of a complex tone. He gave a mapping of percepts ranging from "fused pitch shift" to perception of "dissonant pairs" with other perceptual correlates being "fused roughness", "rolling" or "segregation", for different magnitudes of mistuning. However, this mapping was only given as a function of duration, for mistuning of the **4th** harmonic of 200 Hz. Other harmonics were not investigated.

The present experiment also aimed to provide a mapping of percepts associated with "mistuning" of harmonics of a complex tone for different magnitudes and spectral locations of deviation. Such a mapping should ideally reveal the differential contributions of different components to changes in perceived pitch, timbre and fusion.

In acknowledgment of the fact all these percepts may be associated with inharmonic stimuli, the double-layered task assigned in experiment 2 was used again. This task allowed subjects to report all of these percepts *within a single trial.* The experimental variables included the number of the harmonic being mistuned, the magnitude of mistuning, and the type of frequency change being compared within a block (linear or ratio).

5.2 Stimuli:

Two-tone sequences were used as stimuli, in which the first tone was always set to be a "standard" complex with the first 10 harmonics of either 200 or 400 Hz. The second tone differed from the first, in that the frequency of one of its components was displaced upward or downward from the corresponding harmonic frequency by some amount. This displacement in frequency was either linear or proportional.

"Linear" shifts imply that different harmonics were compared for the same absolute amount of frequency change specified by the same number in Hertz. For the 200-Hz standard, the linear shifts examined for different harmonics were $\pm$ 2, 4, 8, 16, 32 and 64 Hz. For the 400-Hz standard, the linear shifts examined were $\pm$ 4, 8, 16, 32, 64 and 128 Hz.

"Proportional" or "ratio" shifts imply that different harmonics were compared for the same relative amount of frequency change, specified by the same ratio relative to the frequency of the corresponding harmonic component in the standard tone. For both the 200-Hz and 400-Hz standards, the values of proportional shifts examined were $\pm$ 1%, 2%, 4%, 8%, 16% and 32%.

All components in both tones were specified to have the same amplitude and initial phase. Both tones were 100 msec in duration, with 10 msec rise and fall times. A 300-msec silent interval separated them in the sequence. The sequence was repeated twice per trial, after an interval of 1600 msec.

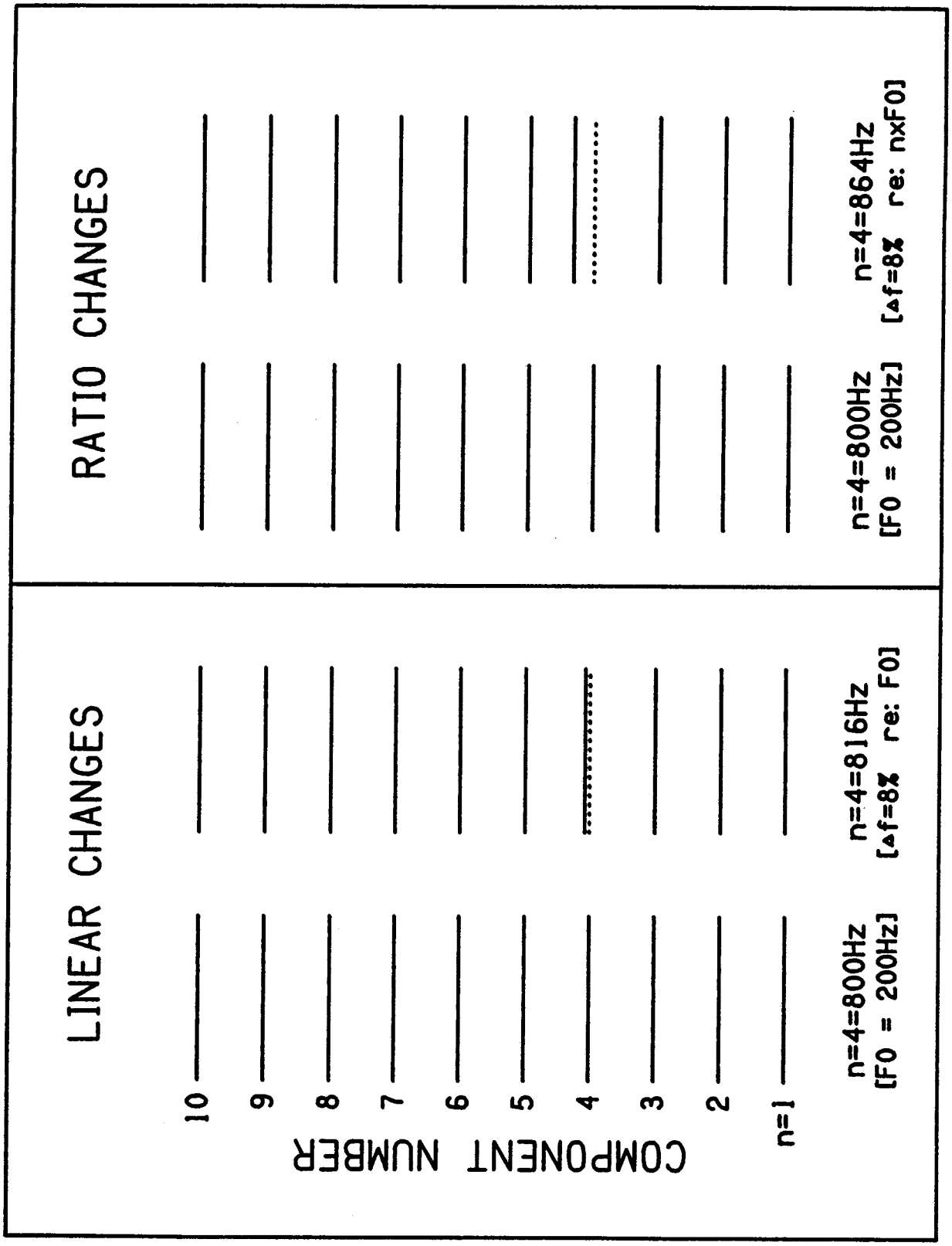
Figure 5.1 illustrates the difference between stimuli in which a single spectral component of the second sound is shifted in frequency by a linear amount, or a ratio relative to the frequency of the standard harmonic component in the first tone.

5.2.1 Rationale for using two types of frequency changes

For changes in frequency of a single component, there is no basic difference between ratio and linear changes as there was for the stimuli of experiment 2 in which all components were changed. In the latter , the ratio changes led to preservation of harmonicity, while linear changes rendered the complex inharmonic. In the present experiment, both ratio and linear changes resulted in inharmonicity.

The rationale for using these two types of change was basically to provide **different magnitudes** of frequency difference. Linear shifts displaced the mistuned component to proximal positions within the span of the inter-component spacing frequency, while proportional shifts were of greater magnitude and often led to the mistuned component being displaced to positions beyond the span of the inter-component spacing. The new position of the mistuned component would have different perceptual consequences, depending on factors such as frequency resolution, and the mutual interference of components within the same critical band.

Figure. 5.1 (next page) Schematic representation of stimuli used in experiment 3. The vertical dimension represents frequency, and the horizontal dimension represents time (not to scale). Each tone comprised 10 components. The second tone could have a single component displaced either by some absolute Δf Hz, or by some proportion relative to the harmonic frequency of the standard first tone.


LINEAR CHANGES
RATIO CHANGES
COMPONENT NUMBER
10
9
8
7
6
5
4
3
2
n=1
n=4=800Hz
[F0 = 200Hz]
n=4=816Hz
[Δf=8% re: F0]
n=4=800Hz
[F0 = 200Hz]
n=4=864Hz
[Δf=8% re: nxF0]

The displacement would also change the spectral density of the stimulus in different ways. While the spectral region occupied by the displaced component would become more dense, spectral "holes" of different magnitude would result in the vicinity of the original harmonic position. Since spectral spacing is an important parameter in pitch determination and the spectrum in general is a bearer of timbre, these changes in density could be construed as changes in pitch, or timbre or both of these percepts simultaneously.

In addition to providing the basic inharmonicity cue as a potential cue for fission of a complex into multiple sources (see section 2.4.1), the different changes would also provide cues for pitch and timbre change that could vary with the magnitude and location of the frequency change.

5.3 Apparatus:

All stimuli were generated digitally by additive synthesis on a PDP 11/73 clone computer operating at a sampling frequency of 20 kHz. A program entitled PSYACX (Lai et al., 1987) allowed components to be specified in terms of their frequency, phase, amplitude, duration and envelope characteristics. After digital-to-analog conversion, the stimuli were further band-pass filtered between 100 Hz and10 kHz via a brickwall filter (Wavetek-Rockland 751A). The response of the filter rolled off by 115 dB/octave outside the passband. The stimuli were amplified to a comfortable listening level (about 70 dB SPL) using digital

attenuators in conjunction with power amplifiers, and were presented binaurally over headphones (AKG 141) to the ears of a listener seated in a sound-absorptive listening room. The accuracy of the synthesized spectra was verified via a Hanning-window analysis using the Bruel & Kjaer 2033 single-channel spectrum analyzer. A Tektronix dual-trace oscilloscope was used to view the stimulus waveforms and to verify the accuracy of the temporal layout of the stimuli in a sequence.

5.4 Procedure

5.4.1 Subjects:

The seven listeners (CS, GC, JO, LC, LJ, WK and JK) used in experiment 2, also served as subjects for experiment 3. In the experimental condition corresponding to ratio changes made in the frequency of harmonics of F0=400 Hz, however, subject LJ was unable to participate. This change in the number of listeners from seven to six is indicated where needed in the results section.

5.4.2 Task:

As was the case in Experiment 2, the task assigned to listeners was twofold: First, they were required to comment on the degree of "fusion" of the complex tones in the stimulus, and secondly, on any

change perceived in the overall pitch or timbre of the sounds. This was accomplished by asking listeners to make two judgements for every stimulus. The instructions were phrased as follows:

First: "Do the tones appear to "segregate" or "split" into more than one entity? If so, which of the tones splits? 1) first tone, 2) second tone, 3) neither, 4) both? Please press one of the keys labelled '1', '2', '3' or '4' on the computer keyboard to indicate your choice."

Second: "Is the second tone in the pair 1) same, 2) higher in pitch, 3) lower in pitch, 4) different in "something else", 5) different in "something else" and higher, 6) different in "something else" and lower in pitch than the first tone ? Please press one of the keys labelled '1', '2', '3', '4', '5' or '6' on the computer keyboard to indicate your choice."

To help listeners remember the designated labels for the task, an illustration of the type used in experiment 2 (shown in **figure 5.2**) was provided during each session.

The first tone in a sequence was not expected to "split" as it remained harmonic and unchanged across sequences and trials. Only the second tone in a sequence was made "different" by mistuning of a harmonic. Thus, the responses '1' and 'B' would not be expected to be used. The use of these labels would be equivalent to "false alarms" (as observed in signal detection tasks (Green and Swets, 1974).

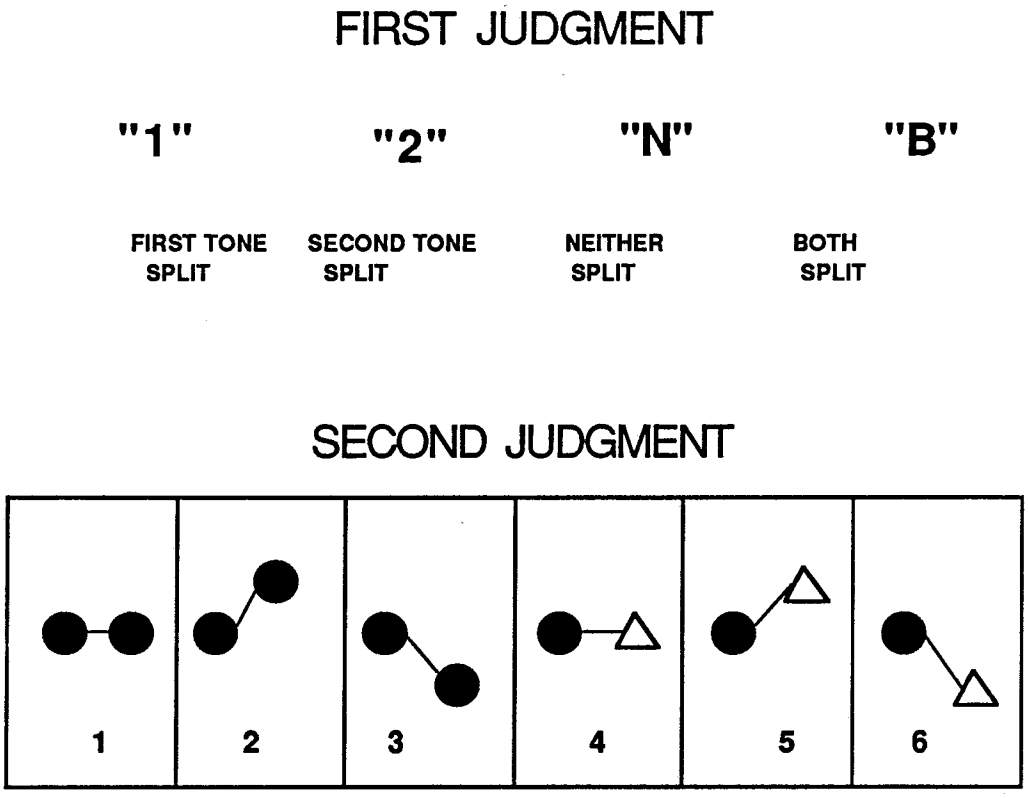


Figure. 5.2 Same as figure 4.2; illustrating the double-task assigned in experiment 3. The first judgment pertained to perceived splitting or fusion of the tones. The four response keys corresponded to perceived splitting of 1) the first tone (label '1'), 2) the second tone (label '2'), 3) neither tone ((label 'N'), or 4) both tones (label 'B') The second judgment pertained to perceived differences in the pitch and timbre of the tones as illustrated.

To avoid indiscriminate "guessing" and arbitrary use of these "placebo" response options, listeners were told that all options were not equally represented in blocks of stimuli. They were further told that there were no "right" answers for the tasks and that they should respond as faithfully as they could, using the labels that seemed most appropriate on each trial, unbiased by the frequency of use of other labels in the block.

5.4.3 Stimulus presentation:

A typical experimental session lasted 2 hours and was divided into "blocks" of trials. Stimuli with different magnitudes of frequency change were grouped into blocks based on the number of the harmonic being displaced. Thus, there were 10 such blocks for each harmonic being mistuned. Each block comprised 13 stimuli of differing magnitude of frequency change. Each of the thirteen stimuli were replicated 4 times within a block, the number of trials within a block thus being 4 X13=52. This type of "blocking" of stimuli was done for both types of change (linear/ratio) and for both values of standard F0 (200/400 Hz). The order of presentation of stimuli was randomized within a block. The order of "blocks" was also randomized across subjects.

After presentation of the 2 repetitions of the stimulus, the listener was given 3.5 seconds in which to make a response. A "response"

comprised the sequential pressing of 2 keys on a keyboard to indicate the choice of labels for the 2 judgements required. Thus, the listener selected one of the four keys 1, 2, 3 or 4 for the first judgement pertaining to "splitting" of the tones, and one of the six keys 1, 2, 3, 4, 5 or 6 for the second judgement pertaining to the pitch and timbre of the tones.

Upon elapse of the 3.5 second interval, a written message on the computer screen before the listener prompted the listener to make a response. A trial-by-trial record of the stimuli presented and the corresponding responses was maintained by the computer.

5.5 Schemes of data representation

The number of times particular response labels were selected to describe the stimulus presented across all 4 replications constituted the raw data. Thus, if for the first judgment a listener chose label 2 to indicate the perceived splitting of the second tone of a stimulus sequence, on 3 trials out of the 4 replications of a sequence, the response label 2 received a score of 3/4 or 75%. For each listener, two response "matrices" were available:

For the first judgment, a (4X13) matrix of responses giving proportions of the **4** possible responses (expressed as % trials) for each of the 13 stimuli presented in a block.

For the second judgment, a (6X13) matrix of responses giving proportions of the **6** possible responses (expressed as % trials) for each of

the 13 stimuli presented in a block

These data have been processed further in 2 different ways for display as was done for experiment 2:

i) To provide a quantifiable measure of the distribution of responses for a particular stimulus configuration, the **proportions of use** of individual labels have been averaged across subjects, while maintaining differences across response categories. To this end, the 7 pairs of response matrices (one pair from each subject) were averaged to give a pair of mean matrices of responses. Each cell of a mean matrix gave the average number of times (in %) that a particular label (1,2,3, or 4 for the first task, or 1,2,3,4,5, or 6 for the second task) was used in response to a particular sequence with a frequency change made in a particular harmonic.

The pattern of response selection is displayed graphically in **figures** 5.3, 5.5, 5.7, 5.9, 5.11, 5.13, 5.15 and 5.17 for the different stimulus conditions. Each frame in these figures represents the number of the harmonic that was being "mistuned". The abscissae give the magnitude of change (in Hertz for the linear changes and in % (re: nXF0) for the proportional shifts), while the ordinates give the proportions of use of labels in terms of percentage of total numer of trials over all replications of the same stimulus.

The data symbols refer to the response label selected. For the first judgment, the symbols '1', '2', 'N' and 'B' are being used to imply the 4

options available for the first judgment. Thus, '1' implies perceived splitting of the first tone, '2' implies perceived splitting of the second tone, 'N' implies that neither tone split, and 'B' implies that both tones appeared to split. For the second judgment, the symbols '1', '2', '3', '4', '5' and '6' are used to indicate the perceived pitch and timbre relations between the two sounds as illustrated in **figure 5.2** .

This type of display thus gives an idea of magnitude of use of the labels and is also shows how the response selection changes as stimulus parameters (such as Δf and n) change across the stimulus set.

ii). In order to show **listener agreement or variability**, each of the response matrices was examined for each listener individually and the label used most frequently in response to a particular stimulus was determined for each listener and each stimulus for both tasks. These "labels of choice" for each stimulus have been plotted in terms of the number of listeners who selected them.

For purposes of graphic representation, the display scheme is similar to that used in the previous graphs, with the difference that the ordinate represents the **number of listeners** favoring a particular response label, rather than the proportion of use of that label. As before, the data symbols '1', '2', 'N' and 'B' refer to the response label selected for the first judgment, and the symbols '1', '2', '3', '4', '5' and '6' refer to the response label selected for the second judgment. These numbers should not be confused with the "number of listeners", which is the variable

represented along the ordinate.

Since the number of listeners was seven, the limit of the ordinate is the number 7. In cases where the numbers of listeners for a particular stimulus configuration adds up to a number greater than 7 reflects equivocation across response labels as has been described previously. This second type of display thus gives an idea of distribution of the "dominant" choice of label across listeners.

5.6 Results for linear changes for Standard F0=200 Hz

Figures 5.3 through 5.6 show response selection for stimuli in which linear changes were made in the frequency of the first10 harmonics of 200 Hz. Each "frame" of the displays corresponds to a different harmonic being mistuned. The abscissa for each graph gives the magnitude of change made in the frequency of the harmonic. The ordinate gives either the percentage of trials on which different response labels were selected, averaged across the 7 listeners, or the number of listeners that selected a particular response label. The ordinate cuts the abscissa at Δf=0 Hz, with upward changes in frequency shown on the right, and downward changes showng on the left.

5.6.1 First judgment responses: Average trend

Figure 5.3 shows the average proportion of labels used for the

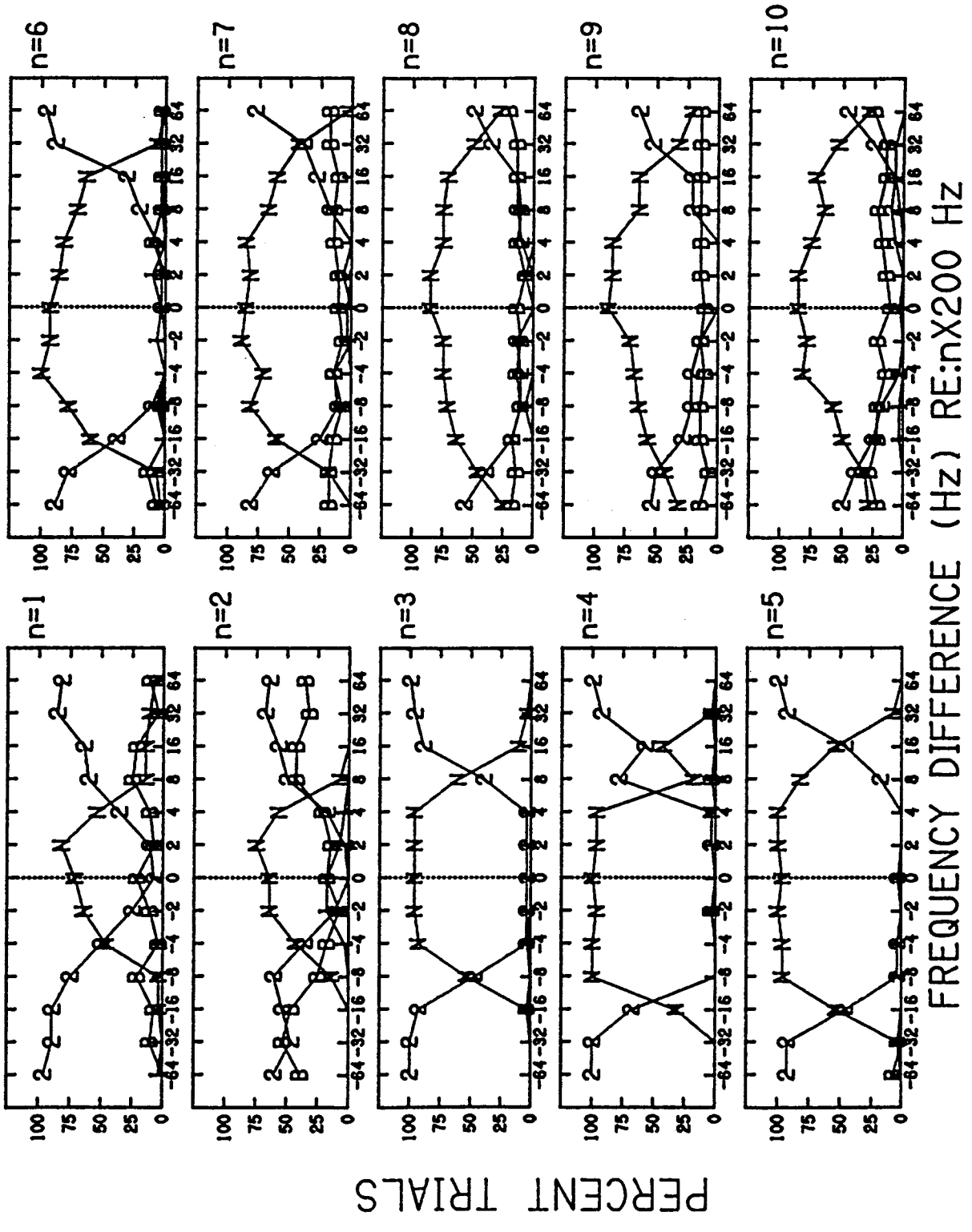
first judgement. The symbols '1', '2', 'N' and 'B' denote the 4 options available for the first layer of the task, i.e. that 1) the first tone split, 2) the second tone split, 3) neither split, 4) both split.

At first glance, it is clear that changes in frequency of high harmonics are less likely to yield segregation than changes in the frequency of lower harmonics. This is indicated by the broadening width of the 'N' curve with increasing harmonic number 'n'.

Also clear from figure 5.3 is the fact that for changes in the frequency of most harmonics, the label '2' (to indicate splitting of the second tone), and the label 'N' (to indicate no splitting) dominate the responses made by listeners. This is in keeping with *a priori* expectations since only the second tone was expected to split as described earlier in section 5.4.2.

For ***n=1***, a 4 to 8 Hz change seems to be adequate to cause perceived splitting of the second tone. For ***n=2***, an 8 Hz change is enough. However, a large number of 'B' responses are also seen, implying that some subjects perceived the first, "standard" tone as being split as well, despite there being no change in the frequency of its components ! For ***n=3 through 6***, this apparently anomalous percept of both tones splitting goes away and the segregation threshold for the second tone increases from 16 to 32 Hz. For ***n=7 through10***, the segregation threshold for the second tone goes up from 32 to 64 Hz and an increase in the number of 'B' responses is also notable.

Figure 5.3 (next page) Data showing label choice (averaged over 7 listeners) for the **first judgment** for sequences with **linear** changes in the frequency of components of magnitude as shown along the abscissa (**re: nX200 Hz**). Proportion of trials on which the labels were used is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies splitting of the first tone, '2' implies splitting of second tone, 'N' implies perceived fusion , 'B' implies that both tones split.


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
PERCENT TRIALS
FREQUENCY DIFFERENCE (Hz) RE:nX200 Hz

5.6.2 Digression: Perceived splitting of both tones despite no changes in the frequency of the first tone.

The use of the 'B' labels was not expected at the design stage of these experiments. Over the course of stimulus production and listening however, it was observed that a "split" component could indeed often be heard in the first, unchanged tone. This type of "analytic" listening was however, not consistently observed across listeners. Some listeners were more "analytic" than others. Furthermore, the splitting of the first tone appeared to be contingent upon the splitting of the second, i.e. the first tone was rarely reported as being split if the second tone remained fused. This confounding indicates that the split component of the first tone is "heard out" only in the context of the changed component of the second tone. Thus an ascent or descent in pitch is heard in the form of a melodic interval against the background of the other components of the two tones. The audibility of a split component in the second tone *retrospectively* makes a component in the first tone audible as well. This type of **"relative" perception** is the hallmark of grouping processes that enable perceptual organization and is discussed in greater detail in Chapter 6.

5.6.3 First Judgment Responses: Listener Variability

Figure 5.4 shows the selection of dominant labels in terms of the number of listeners who selected them. A high percentage of use of a particular label in figure 5.3 is seen to corrsepond to the high number of listeners selecting that label shown in figure 5.4. For n= 3 through 7, the listeners were in excellent agreement with each other, unanimously labelling a sequence as being fused or split dependent on the magnitude of frequency change. For the high harmonics 8, 9 and 10 and the lowest harmonics 1 and 2, there is some variability in response choice. This variability is manifested in the increased use of label 'B'. For n=2 in particular, listeners appeared to be split (no pun intended) in their choice of label '2' to indicate splitting of the second tone, or 'B' to indicate splitting of both tones. The "analytic" mode of listening appears to have been enhanced for changes in frequency of this harmonic and its unchanged counterpart in the first tone.

5.6.4 Second Judgment Responses: Average Trend

Figure 5.5 shows the proportion of responses made for the second judgement for the same stimuli as those represented in figures 5.3 and 5.4. The numbers '1', '2', '3', '4', '5' and '6' denote the six choices available for the second layer of the task as described earlier in Experiment 1 and in section 5.4.2.

Figure 5.4 (next page) Distribution of dominant response labels across listeners for the **first judgment** for stimuli with **linear** changes in a single component **re:nX200 Hz.** Magnitude of change is shown along the abscissa and the number of listeners using different labels is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies splitting of the first tone, '2' implies splitting of second tone, 'N' implies perceived fusion , 'B' implies that both tones split.

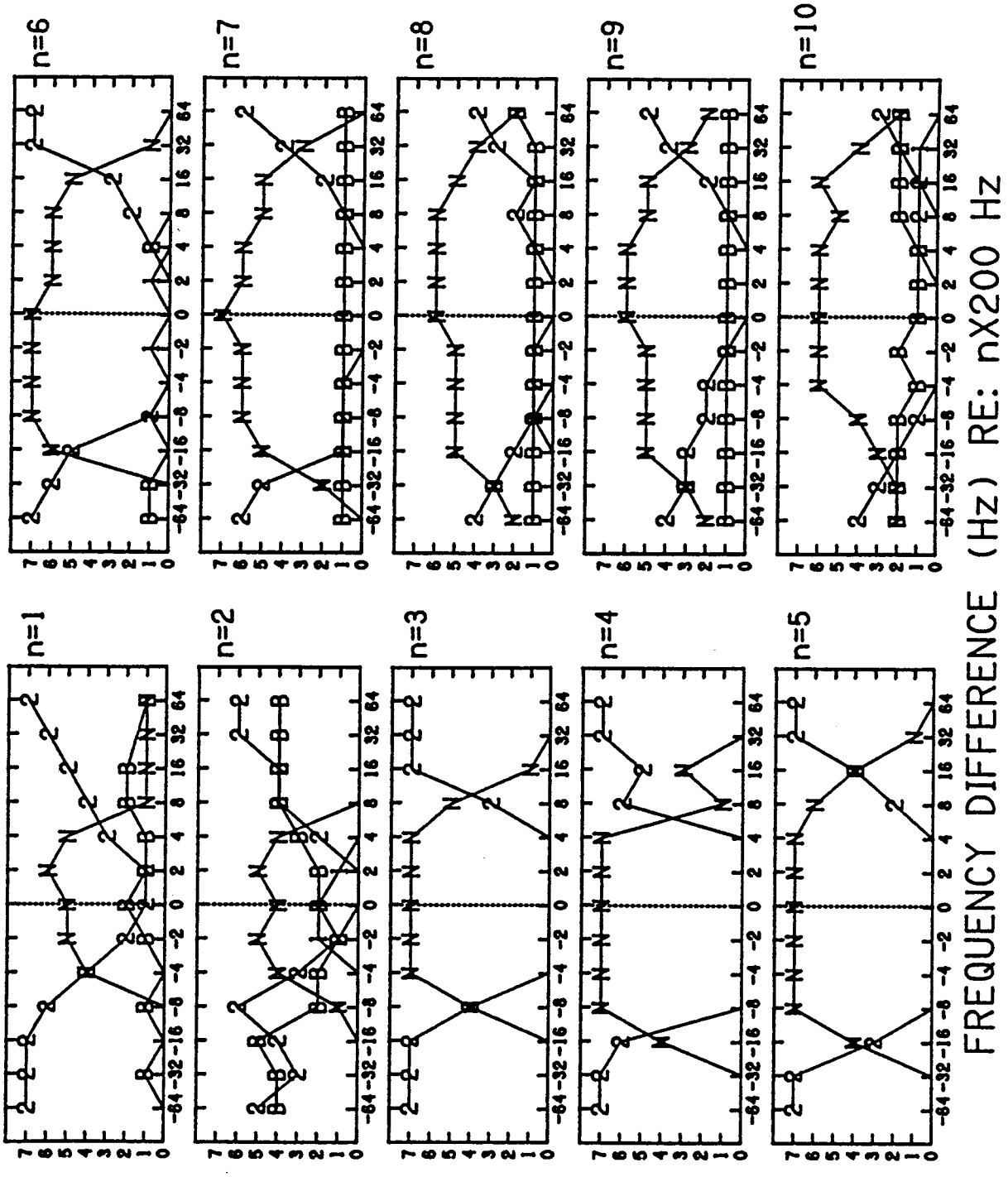
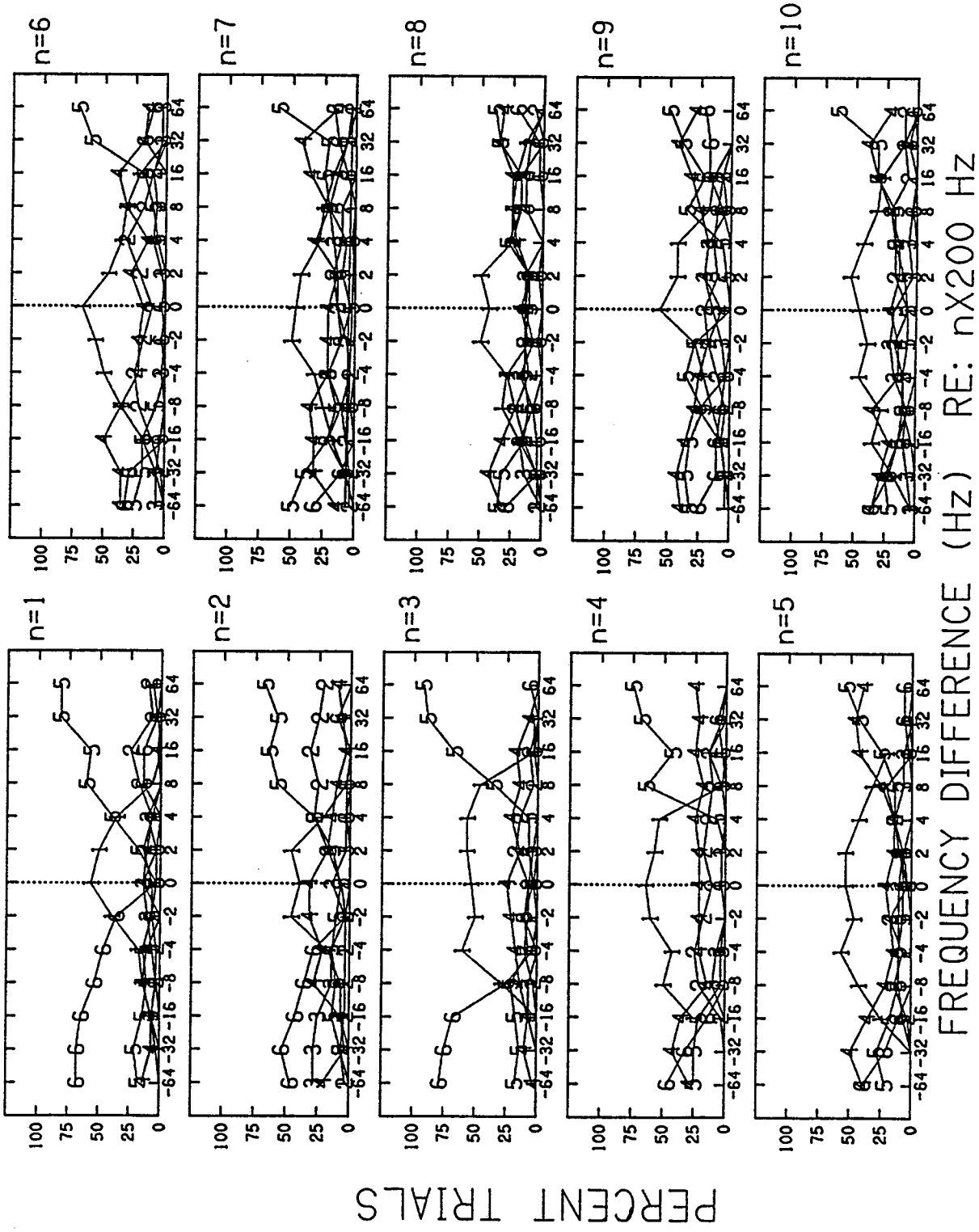

n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
-64 -32 -16 -8 -4 -2 0 2 4 8 16 32 64
0 1 2 3 4 5 6 7
FREQUENCY DIFFERENCE (Hz) RE: nX200 Hz
NUMBER OF LISTENERS (OUT OF 7)

Figure 5.5 (next page) Data showing label choice (averaged over 7 listeners) for the **second judgment** for sequences with **linear** changes in the frequency of components of magnitude as shown along the abscissa (**re: nX200 Hz**). Proportion of trials on which the labels were used is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies no change, label '2' implies a perceived rise in pitch, label '3', a fall in pitch, label '4', a change in "something else", label '5', a rise in pitch and a change in "something else", label '6', a fall in pitch and a change in "something else".


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
PERCENT TRIALS
FREQUENCY DIFFERENCE (Hz) RE: nX200 Hz

For low harmonic numbers (1 through 3), negative and positive changes in frequency are accompanied by perceived changes in both pitch and "something else", as indicated by the use of labels '6' and '5' respectively. On comparison with the labels chosen for the first judgment (Figure 5.3), it seems that the pitch change is correlated with the perception of splitting of the second sound (label '2'). Thus, a change in the frequency of these low components is perceived as a change in pitch in addition to enabling them to be heard out as separate entities.

It is not within the scope of the design of the second task to disambiguate exactly **which pitch** was being compared across the sequence; whether the "virtual", overall *gestalt* pitch of the two complexes, or the "spectral pitch" of the separated component and its counterpart, or perhaps both spectral and virtual pitch. As discussed in the section on pitch in chapter 2, spectral pitch is contributory to the pitch of the complex.

The experiments of Moore et al. (1985 a) showed that changes in the frequency of low harmonics 1 through 6 contributed to a "pitch shift" for the complex as a whole. Their data verify the "dominance" phenomenon studied by Plomp (1967) and Ritsma (1967). However, the latter two investigators showed the "optimal" harmonics for pitch to be n=3 and 4. The data for the second judgment shown in figure 5.5 however, show a greater ability to follow pitch change for n=1 and 2.

It could be that the experimental design used here sets up a

competition between two modes of listening. The separability of a component from a complex dictates the "analytic" mode. The perception of overall pitch on the other hand, dictates "synthetic", holistic listening.

The correlation of the "split" judgments with the perception of pitch changes points toward analytic listening where the pitch of the pure tone is heard out separately. However, the fact that the direction of pitch change is reported accurately for changes in these low harmonics presents a dilemma. If a component was being "heard out" and its emergent pitch was taken to indicate a change in pitch *across* the two tones of the sequence, then one may expect listeners to simply report an ascending pitch ('5') since the component frequency is typically higher than the F0 of the remaining complex (except for descents in frequency of n=1). The use of label '6 to indicate descents in pitch implies however that listeners were sensitive to the direction of change in frequency of a harmonic. Even a segregated component (n=2, 3) with a frequency higher than the reference F0 (200 Hz) was able to convey the correct direction of change of frequency.

There are two plausible hypotheses to resolve this dilemma. The first is that of *"enhanced" analytic listening*. According to this suggestion, the audibility of the changed component in the second sound retrospectively enables the corresponding unchanged component in the first sound to be heard out by virtue of its frequency proximity that facilitates matching of spectral pitch. The influence of frequency proximity on perceptual grouping is a well-known phenomenon

(Bregman and Campbell, 1971) and would support the idea of the "streaming" together of the partial components being changed, leaving the rest of the complex in a separate perceptual group.

Plomp (1964) and von Helmholtz (1856/1954) before him, also used "probe" tones to tune the attention of a listener to a particular frequency region to facilitate analytic listening. The results of the present experiment indicate that this can be done retrospectively as well. With the context of the unchanged component thereby provided in the first sound, the tracking of frequency change would be more accurate and downward shifts would lead to descents in perceived pitch.

The second hypothesis would be one of *"synthetic" listening* that incorporates the change in frequency of a partial into the pitch extraction process. The shift in the frequency of a component would then influence its contribution to the overal pitch. For "dominant" lower components, this contribution would be manifested in the direction of the perceived pitch change. The use of the label '6' may thus be due to perceived changes in overall (versus "spectral") pitch.

Another ambiguity in the data for the second judgment responses is the meaning of the change in "something else". This term, while considered synonymous with a change in timbre, may be construed in 2 different ways in this particular experiment. Thus, a listener may report the "splitiing" of a sound as a change in "something else". For lower components where pitch change is reported along with a change in "something else", this may indeed be the case. For cases where the label

'4' is used, there is less ambiguity. For such sequences, there is no confounding between "splitting", "pitch change" and "something else". **A measure of timbre change in this experiment may thus be taken to be the proportion of use of label '4'.**

For n=2, figure 5.5 shows that the labels '3' and '2' were also selected on about a quarter of the trials for downward and upward changes greater than 4 Hz. This implies that the frequency change was occasionally construed as a change in overall pitch of the complex, *without* an accompanying change in "something else".

On comparison with the choice for n=2 for the first judgment, it appears that the sequences that were labelled 'B' to imply splitting of both tones, are the same ones that have been labelled '3' and '2' here. These stimuli thus appear to be heard as 2 sounds that are "split" in a *parallel* fashion (like a chord sequence). Thus, despite the splitting, no change in "something else" is reported across the two sounds. Instead, only the net direction of the perceived change in pitch is reported.

For higher harmonic numbers (4 through10), an upward change in frequency appears to be heard as a change in pitch and a change "something else" more saliently than a downward change, as indicated by a higher proportion of '5' responses for changes ranging from +16 to +64 Hz. The proportion of '6' responses for the same magnitude shift downwards is not as dominant. The variability in response is also increased for higher harmonics, as indicated by the increased use of other labels. The label '4' denoting a change in something other than pitch is

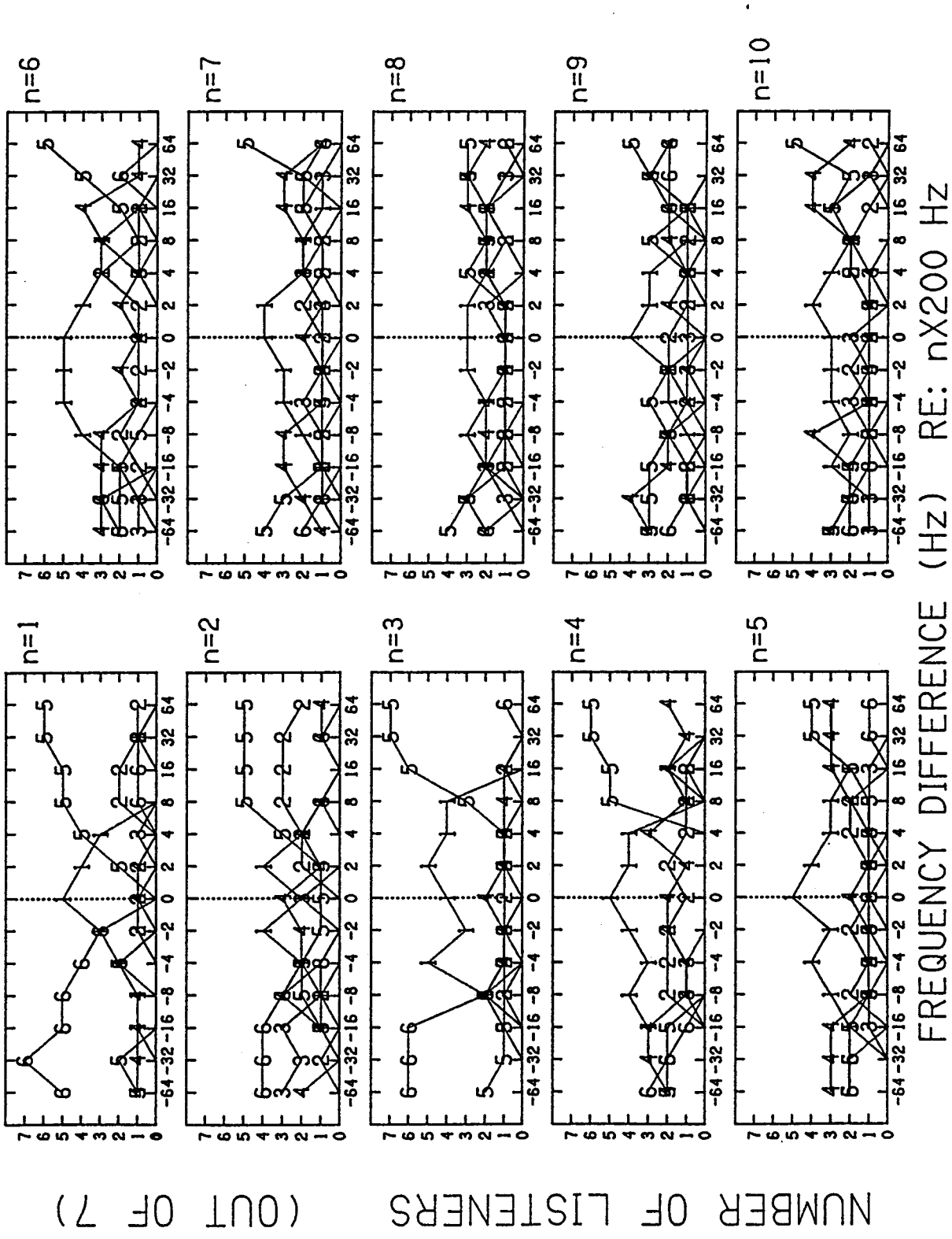
chosen more and more frequently for n=5 through 9. For n=10, '1' is the dominant label implying that changes in frequency ranging from 2 through 32 Hz were not perceived and the two tones appeared to be identical.

A point of interest is the fact that the label '5' is used by listeners to indicate a rise in pitch for some stimuli in which the frequency shift is infact negative. The proportion of trials on which this occurs is seen to increase from less than 25 % for changes greater than -16 Hz in n=1, to 50% for a -64 Hz change in n=7. Conversely, for some stimuli, the label '6' indicating a downward pitch change is also used, despite the frequency change being in the upward direction. This can be seen for +32 and +64 Hz shifts in n=6 through 9. This curious reversal of label choice may be based on comparative "tracking" of the frequency of an adjacent upper or lower harmonic in the standard tone to the shifted component in the second tone, rather than on a comparison based on tracking of the correct harmonic.

5.6.5 Second Judgment Responses: Listener Variability

Figure 5.6 shows that listeners were in good agreement with each other for n=1 and 3. For n=2 and n≥5, the ambivalence reflected in the multiple label choice seen in figure 5.5. is borne out by the divergence of listeners in selecting labels. All the types of modes described in the previous section are manifested for these harmonics.

Figure 5.6 (next page) Distribution of dominant response labels across listeners for the **second judgment** for stimuli with **linear** changes in a single component **re:nX200 Hz.** Magnitude of change is shown along the abscissa and the number of listeners using different labels is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies no change, label '2' implies a perceived rise in pitch, label '3', a fall in pitch, label '4', a change in "something else", label '5', a rise in pitch and a change in "something else", label '6', a fall in pitch and a change in "something else".


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
-64 -32 -16 -8 -4 -2 0 2 4 8 16 32 64
0 1 2 3 4 5 6 7
FREQUENCY DIFFERENCE (Hz) RE: nX200 Hz
NUMBER OF LISTENERS (OUT OF 7)

Some listeners (2/7) appear to be able to track the "true" direction of frequency change by reporting descents in pitch ('6'). About half the listeners (3/7) report timbre changes as manifsested by use of label '4' for the higher harmonics. For the highest increment in frequency (64 Hz), the listeners are more unanimous in their response, using label '5' to report an ascent in pitch and a change in "somethng else".

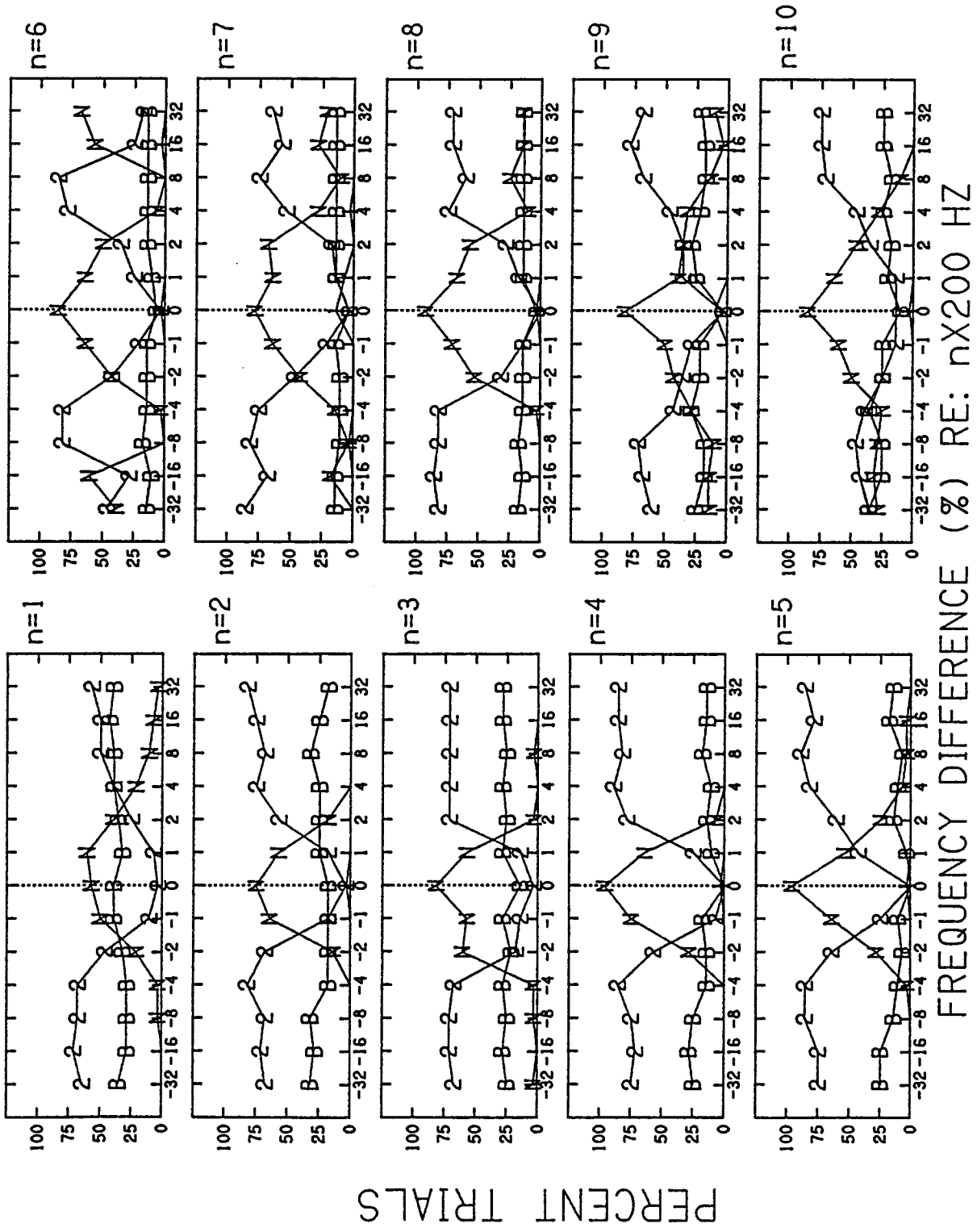
5.7 Results for ratio changes for standard F0=200 Hz

5.7.1 First Judgment Responses: Average Trend

Average data for proportional shifts in the frequency of the 10 components for the 200 Hz complex are given in figure 5.7 for the first judgment.

The obvious difference in comparing linear versus ratio changes in **figures 5.3 and 5.7** is the greater degree of segregation for ratio changes in the frequency of components, than seen for linear changes. The reason for this difference seems to be straightforward, given that ratio changes yield a greater magnitude of frequency change in absolute terms. For higher harmonics, the same ratio will yield even higher magnitudes of actual frequency change.

Figure 5.7 (next page) Data showing label choice (averaged over 7 listeners) for the **first judgment** for sequences with **ratio** changes in the frequency of components of magnitude as shown along the abscissa (**re: nX200 Hz**). Proportion of trials on which the labels were used is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies splitting of the first tone, '2' implies splitting of second tone, 'N' implies perceived fusion , 'B' implies that both tones split.

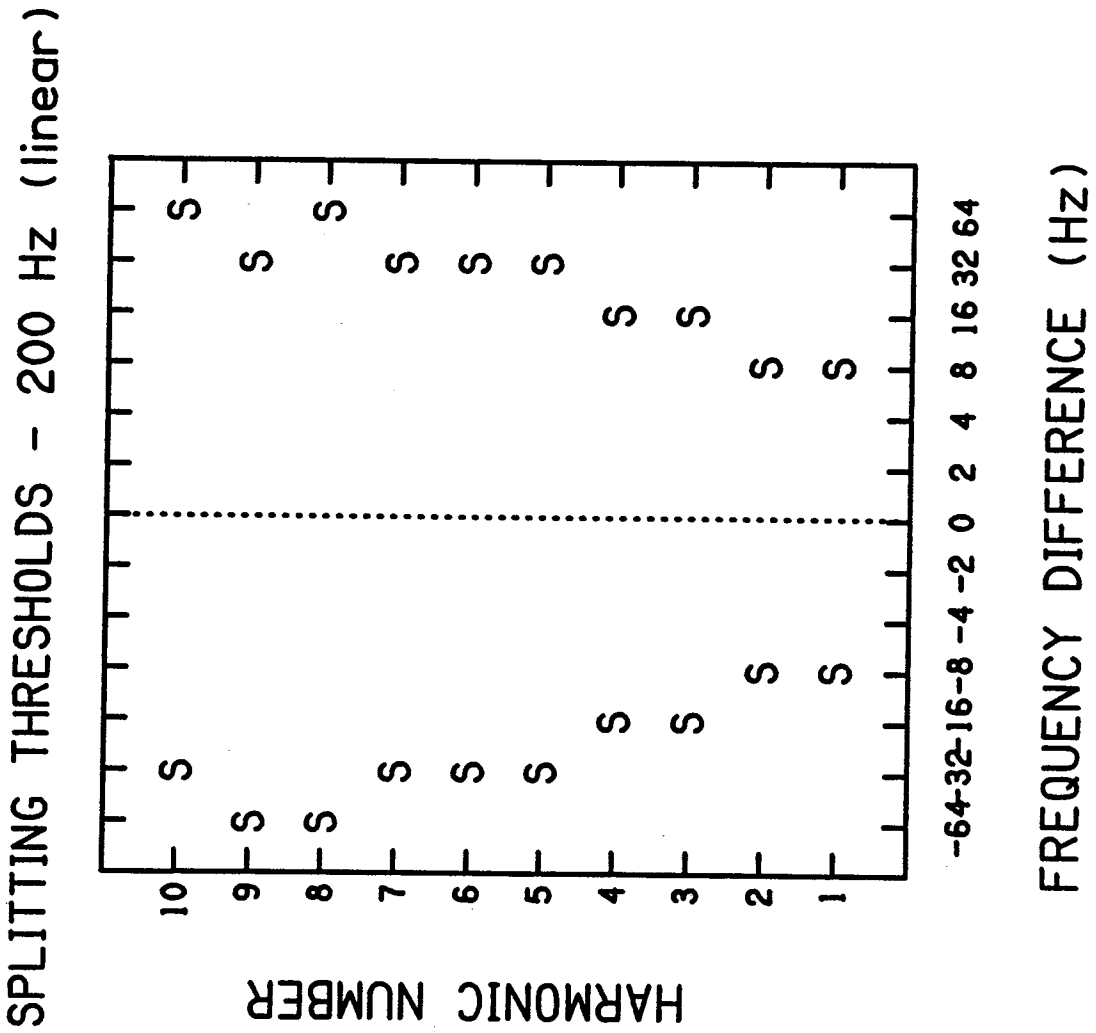

n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
PERCENT TRIALS
FREQUENCY DIFFERENCE (%) RE: nX200 HZ

It may appear surprising at first that the proportion of '2' responses indicating splitting of the changed second tone are smaller for ratio changes than the proportions for linear changes, despite the former changes being larger in absolute frequency terms. This apparent disparity is resolved upon observation of the increased number of 'B' responses for ratio changes, since the splitting of both tones includes the splitting of the second tone by definition. This enigmatic splitting of the first tone was discussed earlier in section 5.6.2 and will be explored further in chapter 6.

5.7.2 Predictability of splitting judgments for ratio changes based on judgments for linear changes

The labels used for ratio changes can be predicted from the "threshold" magnitude of change required for segregation as estimated from the linear-change data . These "threshold" values of Δf are taken to be the points on the abscissae in figure 5.3 where the proportion of use of the label 'N' was exceeded by the combined proportion of use of the "split" '2' and 'B'. The estimated segregation thresholds are plotted in **figure 5.8**, with harmonic number represented along the ordinate, and magnitude of change along the abscissa. The symbol 'S' is being used to imply "splitting".

Figure 5.8 (next page) Threshold value of frequency deviation (shown along the abscissa) required for "splitting" judgments ('2' and 'B' combined), for different harmonics as shown along the ordinate. These "threshold" values were estimated from the label data for the first judgment with linear changes made in components **re:nX200 Hz** (as shown in figure 5.3).


SPLITTING THRESHOLDS - 200 Hz (linear)
HARMONIC NUMBER
10
9
8
7
6
5
4
3
2
1
S
-64 -32 -16 -8 -4 -2 0 2 4 8 16 32 64
FREQUENCY DIFFERENCE (Hz)

From this "threshold" graph, it can be seen that for n=1 and 2, a linear change of about 8 Hz is necessary for segregation. Corresponding ratio changes of 2 to 4% elicit the expected splitting in figure 5.7. For n=3 through 5, figure 5.8 indicates that an 8 to 32 Hz difference in frequency elicits segregation. For ratio changes, the threshold for these harmonics is comparable, corresponding to 2% (=12 Hz) for n=3, 2 to 4% (=16 to 32 Hz) for n=4 and 2% (=20 Hz) for n=5 in figure 5.7.

The data for n=6 **in figure 5.7** appeared rather puzzling. A 4% change seemed to be adequate for eliciting segregation, but for a 16 to 32% change, the dominant response was to report fusion (label 'N').

This interesting observation can be understood when the ratio changes are converted to absolute frequency. Thus, a 4% change for n=6 (1200 Hz) corresponds to a change of 48 Hz. This is above the 32 Hz threshold determined from the linear change data, and thus resulted in a split percept. However, 16% and 32% changes in n=6 correspond to linear shifts of 192 and 384 Hz respectively. The shifted component thus had a frequency of either 1392 or 1008 Hz for 16% and 1584 or 816 Hz for a $\pm$32% change. The first pair of numbers are 8 Hz away from the frequency of the 7th and 5th harmonic respectively, while the second pair is 16 Hz away in frequency from the 8th and 4th harmonics respectively. It could thus be that listeners were making comparisons between the "wrong" harmonic number in the standard first tone, and the changed component in the second sound (as was also noticed in the

experiments of Schouten et al., 1962; Gerson and Goldstein, 1978). If the magnitudes of the linear shift are compared with the linear labelling data for these misconstrued harmonics, the data for ratio changes in n=6 make perfect sense as elaborated below:

The linear "splitting threshold" data (see figure 5.8) predict that n=5 and n=7 would need at least a 32 Hz change in frequency to evoke segregation. Thus, if a listener was comparing these harmonics, rather than n=6, the 8 Hz change was not enough to yield segregation, as observed by the dominant 'N' response for a 16% change in n=6. Similarly, the linear data indicate that a change of 16 Hz in frequency is enough for n=4, but not enough for n=8 to elicit segregation. The 32% change in n=6 thus yielded segregation for a downward change (=16 Hz change in n=4) but not for an upward change (=16 Hz change in n=8).

The data for n=7 through 10 also follow from the linear data which show that a 32 to 64 Hz change or greater is needed for segregation. The 4% ratio change for these harmonics was equivalent to linear changes ranging from 56 Hz for n=7 to 80 Hz for n=10. The second tone of the stimulus is thus seen to segregate for these stimuli.

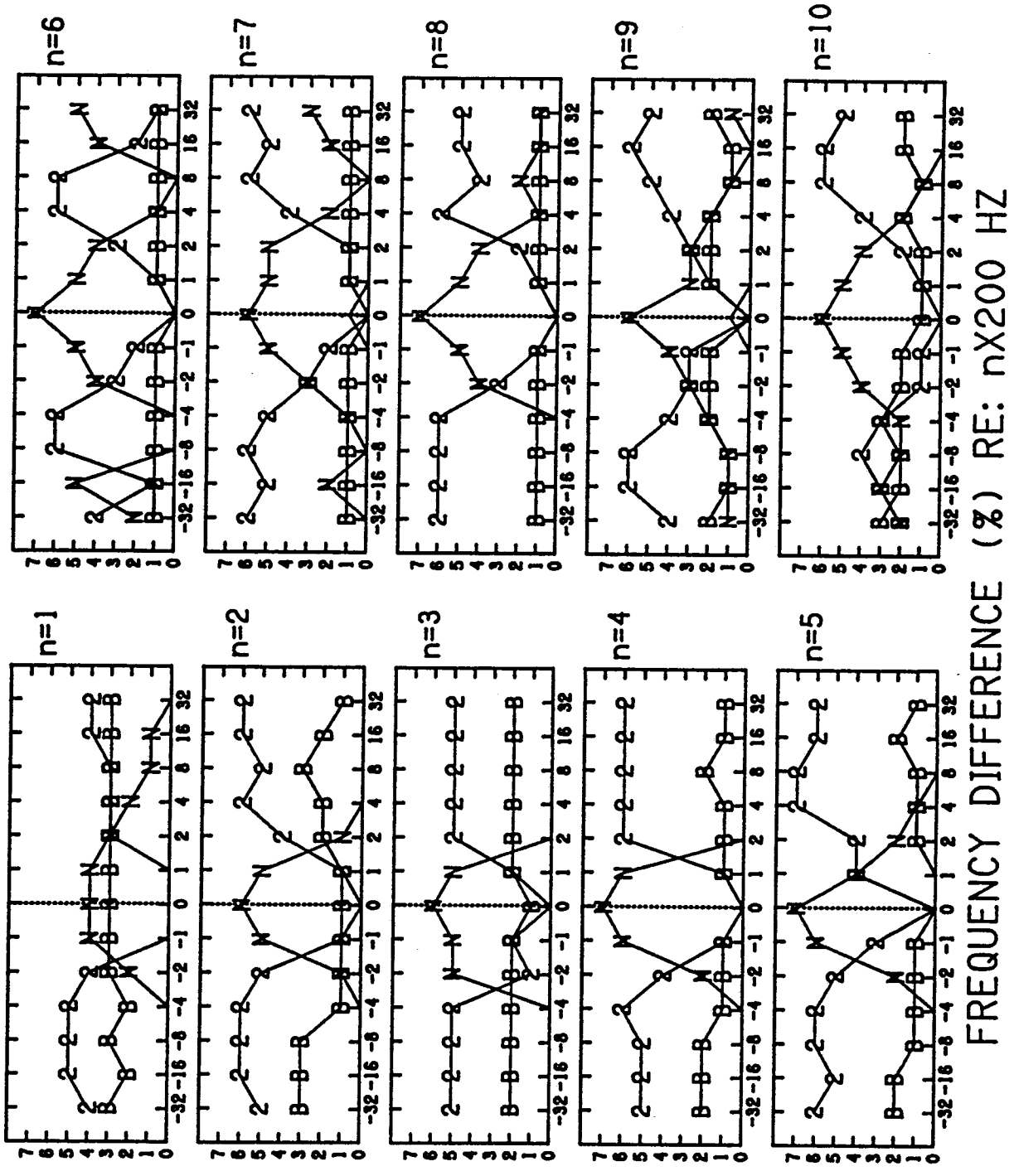
5.7.3 First Judgment Responses: Listener Variability

Figure 5.9, showing response selection for the first judgment in terms of number of listeners, reveals that the 'B' labels in the proportion data (figure 5.7) were contributed by 2 to 4 listeners for n=1 through 4.

One listener (GC) consistently operated in this enhanced "analytic" mode as can be seen by the 'B' label distribution for **all** harmonics, for all magnitudes of frequency change. This listener is a highly-trained musician with a Ph. D. in music. He also teaches ear-training to music students. It seems that he was unable to "fuse" the components into a unitary percept and heard the two tones as chord-like stacks of components. Whether his enhanced analytic-listening ability is a result of "musical training" per se, or general auditory acuity is not clear. The other musician in the group (WK) did not show analytic listening of this type in this experiment. She did diverge from other listeners in experiment 1, however, where she tended to report changes in spectral locus as changes in pitch despite no change in F0.

It may be that *musicians are not a homogeneous group* as sometimes assumed in psychological experimentation. The specific type of training, familiarity with particular instruments and personal auditory acuity may combine to yield different discrimination abilities. There may also be differences *within* the general mode of "analytic" listening. Spectral pitch matching has typically been thought to be a result of listening in the "analytic" mode (Terhardt, 1974). WK's ability to follow spectral locus would then be construed as analytic listening. However, the "hearing out" of components is also an analytic skill. While WK reported hearing out components of the changed second tone along with other listeners, she did not exhibit the "enhanced" retrospective analytic listening manifested by GC.

Figure 5.9 (next page) Distribution of dominant response labels across listeners for the **first judgment** for stimuli with **ratio** changes in a single component **re:nX200 Hz.** Magnitude of change is shown along the abscissa and the number of listeners using different labels is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies splitting of the first tone, '2' implies splitting of second tone, 'N' implies perceived fusion , 'B' implies that both tones split.


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
-32 -16 -8 -4 -2 -1 0 1 2 4 8 16 32
0 1 2 3 4 5 6 7
FREQUENCY DIFFERENCE (%) RE: nX200 HZ
NUMBER OF LISTENERS (OUT OF 7)

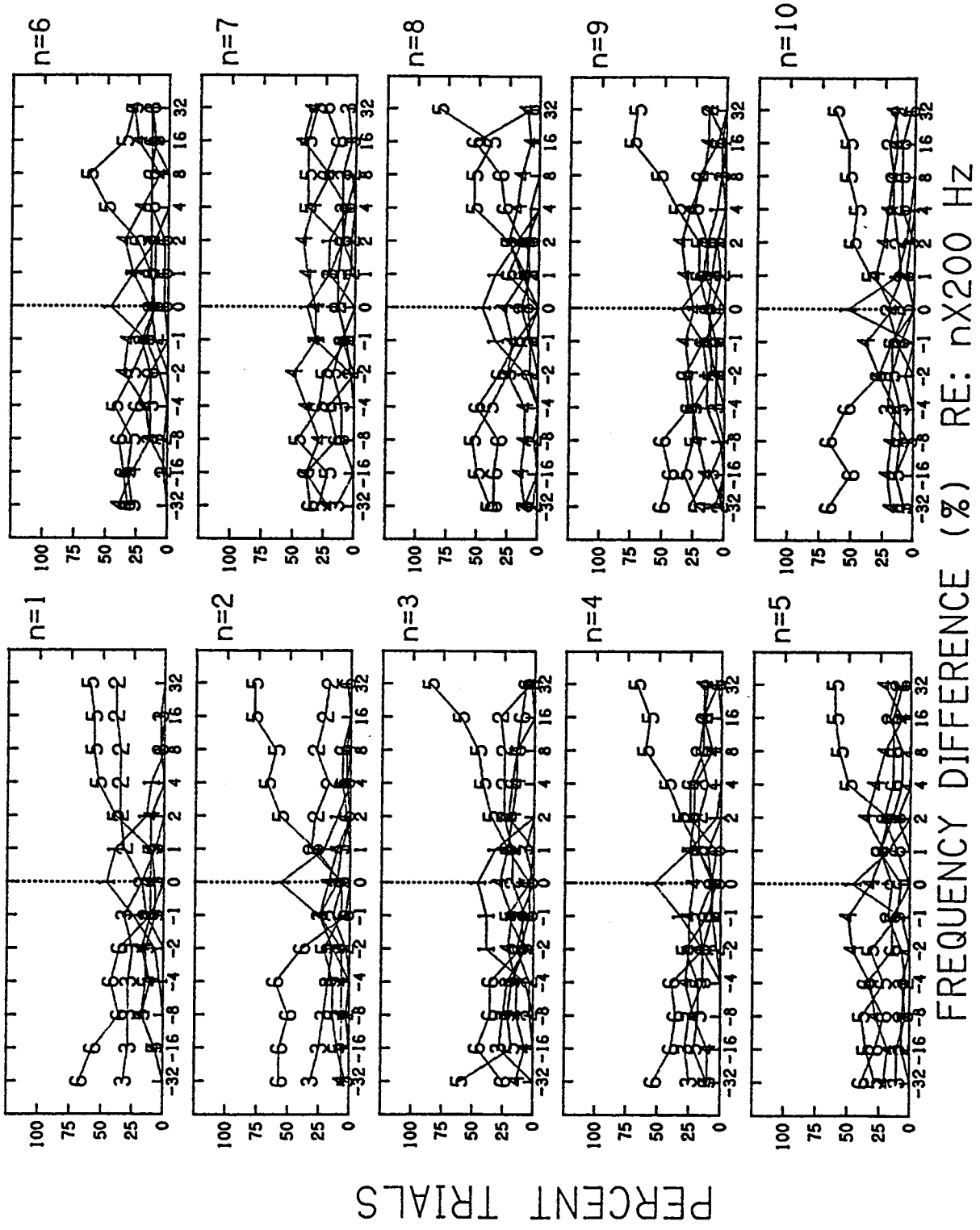
5.7.4 Second Judgment Responses: Average Trend

Figure 5.10 shows the proportion of labels chosen for the second judgement for stimuli employing ratio changes of harmonics. The general trend is similar to that seen in figure 5.5 for linear changes.

Thus, for n=1, 2, 4, 9 and 10, upward and downward shifts in frequency were construed as a rise or fall in pitch accompanied by a change in "something else" as indicated by the dominant use of labels '5' and '6' respectively. For the other harmonics however, some degree of confusion was indicated. While upward shifts in frequency greater than 4% were given the label '5' implying correct judgements of direction of pitch change, downward shifts in frequency evoked variability in response, and in some cases "reversing" of labels as observed for linear changes. Thus for example, a -32% change in n=3 was construed as a rise in pitch (label '5') as were changes of -8 to -32% in n=8.

In terms of actual frequency, the changed component for these stimuli had a frequency of 408 Hz for a -32% shift in n=3, 1472 Hz for a -8% shift in n=8 and 1088 Hz for a -32% shift in n=8. As discussed before, the number of the harmonic being compared could have been misconstrued, and the pitch "tracked" up from an adjacent lower harmonic of the standard first tone, rather than tracked downward from the correct harmonic number in the standard to the lowered component of the second tone.

Figure 5.10 (next page) Data showing label choice (averaged over 7 listeners) for the **second judgment** for sequences with **ratio** changes in the frequency of components of magnitude as shown along the abscissa (**re: nX200 Hz**). Proportion of trials on which the labels were used is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies no change, label '2' implies a perceived rise in pitch, label '3', a fall in pitch, label '4', a change in "something else", label '5', a rise in pitch and a change in "something else", label '6', a fall in pitch and a change in "something else".


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
PERCENT TRIALS
FREQUENCY DIFFERENCE (%) RE: nX200 Hz

5.7.5 Second judgment responses: Listener variability

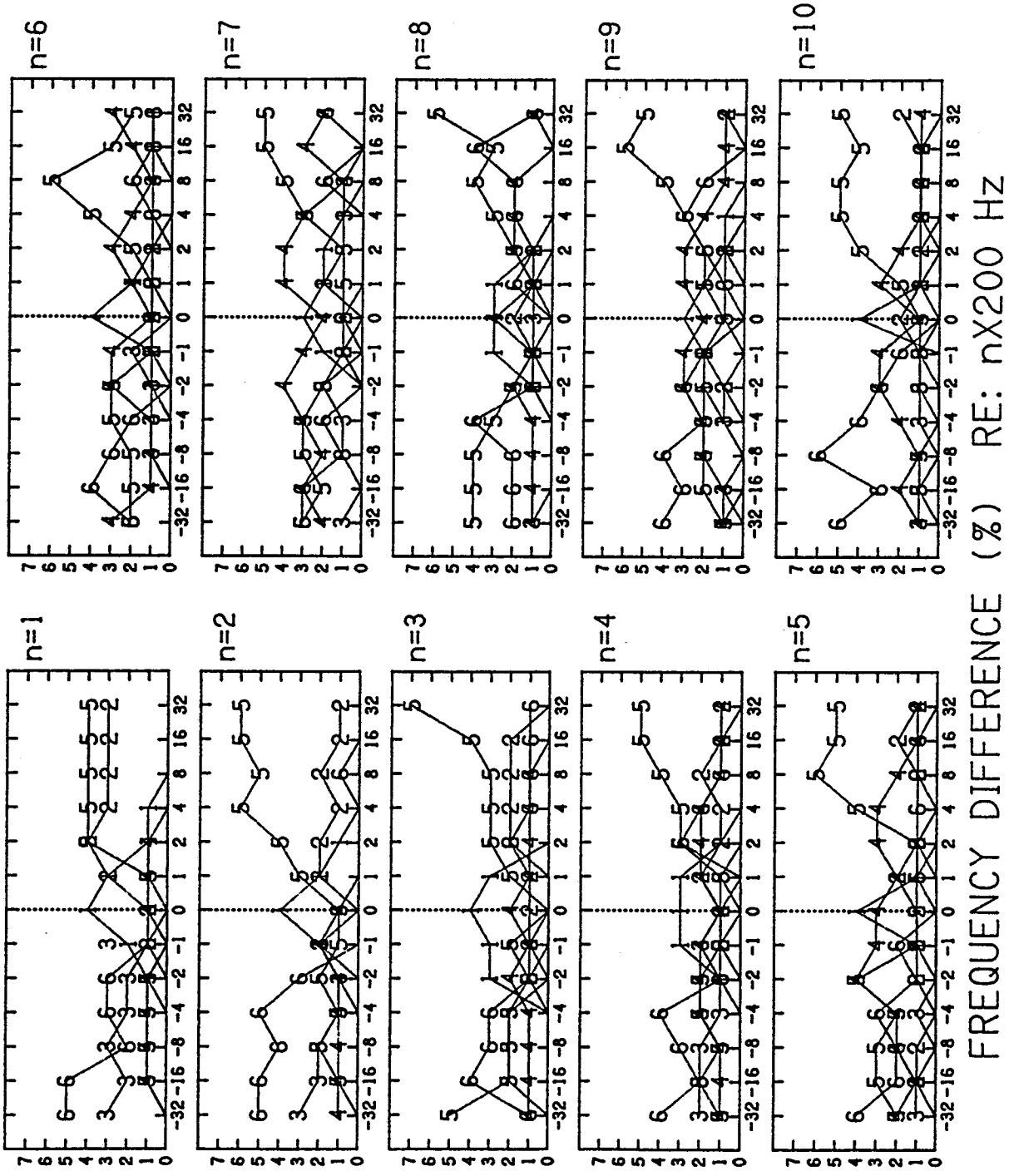
As was the case for linear changes, figure 5.11 indicates that there was considerable variability in response, both across categories, and across listeners, for the second judgment with ratio changes.

It seems that judgments pertaining to splitting and fusion were easier to make because of the clarity of the associated percept of different entities emerging. Judgments about pitch relations and other types of changes in "quality" however were more difficult to make.

Inharmonic stimuli of this type are associated with ambiguous or multiple pitches. The mode of listening further influences the type of pitch the listener is judging: whether the pitch of a single component, or the pitch of the complex as a whole (Gerson and Goldstein, 1978; Schouten et al., 1962). Some listeners may choose to ignore the ensemble of associated potential pitches and report other cues such as "roughness" by using label '4' to indicate a change in something other than pitch.

For low harmonics (n=1 to 3), there was less variability in response but the label distribution for higher harmonics exhibited the whole range of responses described above for the linear case.

Figure 5.11 (next page) Distribution of dominant response labels across listeners for the **second judgment** for stimuli with **ratio** changes in a single component **re:nX200 Hz**. Magnitude of change is shown along the abscissa and the number of listeners using different labels is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies no change, label '2' implies a perceived rise in pitch, label '3', a fall in pitch, label '4', a change in "something else", label '5', a rise in pitch and a change in "something else", label '6', a fall in pitch and a change in "something else".


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
-32 -16 -8 -4 -2 -1 0 1 2 4 8 16 32
0 1 2 3 4 5 6 7
FREQUENCY DIFFERENCE (%) RE: nX200 Hz
NUMBER OF LISTENERS (out of 7)

5.8 Results for linear changes with standard F0=400 Hz

5.8.1 First judgment responses

Labelling data for stimuli in which the F0 of the standard tone was 400 Hz and components of the second tone were shifted linearly in frequency, are shown in figures 5.12 and 5.13 in terms of average proportion of use of labels, and numbers of listeners who selected different labels, respectively.

The major difference between the data for this F0 and for 200 Hz, is the increase in use of the 'B' label for **lower** harmonics (n=1 to 3). For the 200 Hz F0, one listener selected the 'B' label for n=1, for some magnitudes of change (see figure 5.4). For 400 Hz, however, *four out of seven* listeners chose this label for n=1, for all values of Δf, including 0 Hz ! Apparently, analytic listening in the sense of "hearing out" components was highlighted at this spectral region.

Moore et al. (1986) also reported slightly lower thresholds for "hearing out" harmonics of F0=400 Hz, than for those of F0=200 Hz, for low values of n (=1, 2, 3). It **may be** the case, that the absolute frequency of components is a determinant of their vulnerability to segregation, rather than a relative frequency factor such as harmonic number. This is suggested by the observation that the response to n=2 for F0=200 Hz (figure 5.4) was more similar to n=1 for 400 Hz-F0 (figure

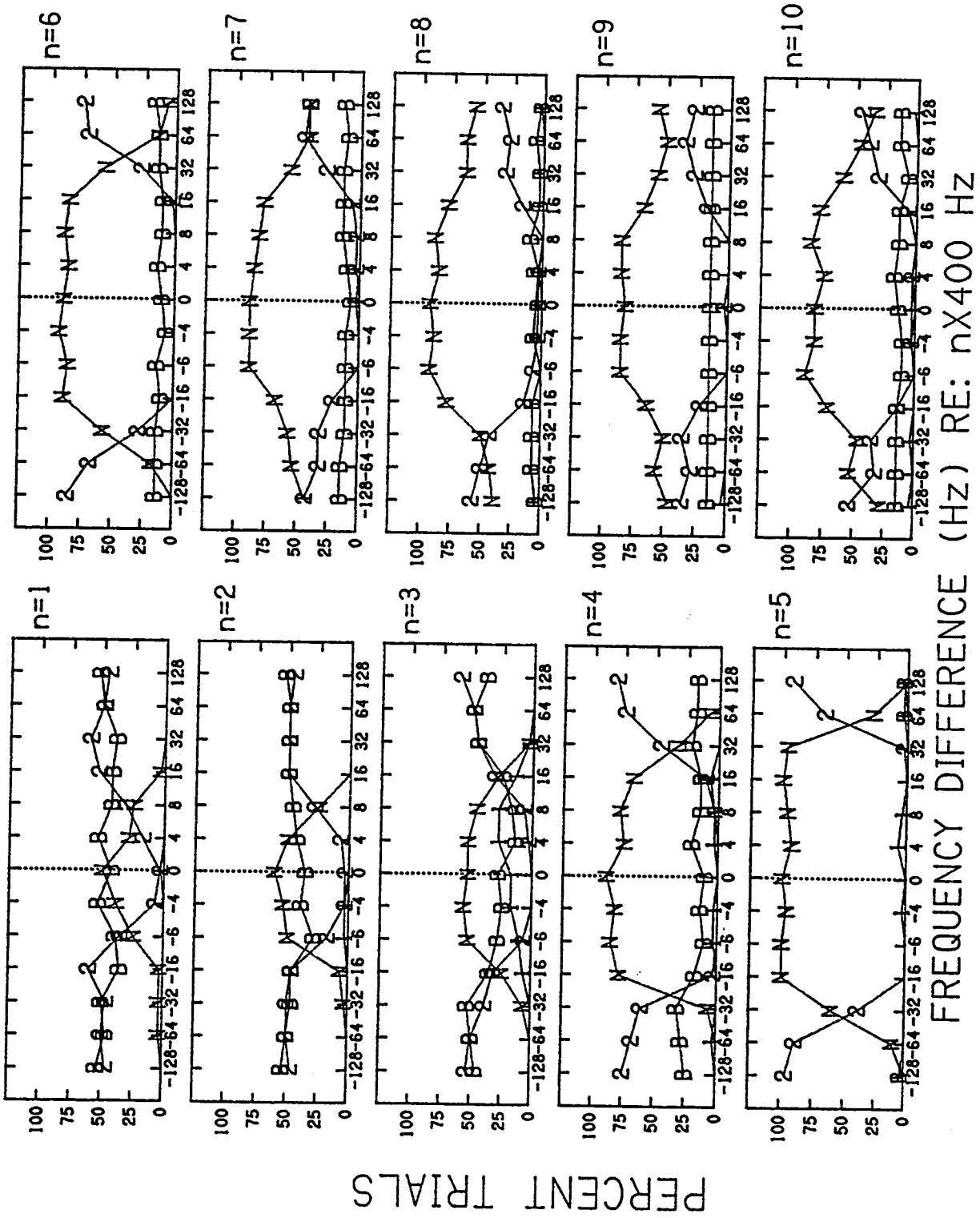
5.13), rather than to n=2. Since, these two harmonics have the same ***frequency*** (400 Hz), similar responses may indicate that absolute frequency is more important than the frequency of the harmonic relative to the fundamental (as defined by 'n').

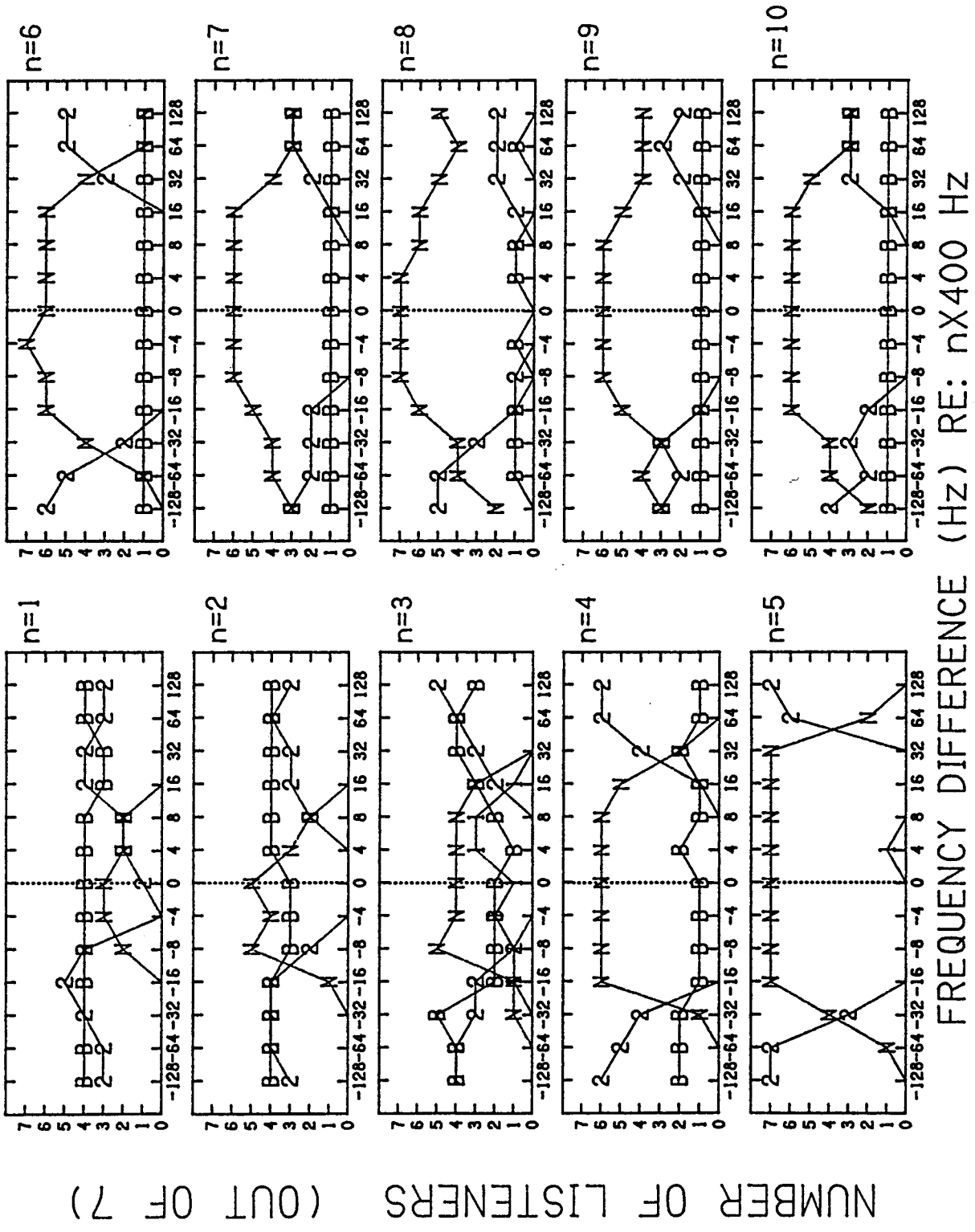
However, this hypothesis does not seem to be borne out for other harmonics. For example, n=2 for F0=400 Hz is 800 Hz, and is equal in frequency to n=4 for F0=200 Hz. The labelling data for the two deviant harmonics are quite different, however. Figure 5.13 shows that 3 to 4 listeners selected label 'B' for n=2, while figure 5.4 shows that for n=4 , all listeners selected label 'N' implying fusion until the deviation exceeded $\pm$ 8 Hz, after which label '2' dominated.

It thus seems that it is not just the magnitude of the frequency change or the standard frequency of the harmonic being changed, that affect judgments of fusion, but it is also the **spectral context** in which these changes are made. The wider spacing of components for the 400-Hz condition, may have been the operative factor facilitating "analytic" listening. Frequency difference of components is indeed an important factor in "hearing out" of components (Plomp and Mimpen, 1967; see section 2.3). If two components are separated by more than the critical bandwidth, they are less likely to interact, and are more amenable to being heard out under appropriate listening conditions.

Figure 5.12 (next page) Data showing label choice (averaged over 7 listeners) for the **first judgment** for sequences with **linear** changes in the frequency of components of magnitude as shown along the abscissa **(re: nX400 Hz)**. Proportion of trials on which the labels were used is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies splitting of the first tone, '2' implies splitting of second tone, 'N' implies perceived fusion , 'B' implies that both tones split.

Figure 5.13 (page 333) Distribution of dominant response labels across listeners for the **first judgment** for stimuli with **linear** changes in a single component **re:nX400 Hz.** Magnitude of change is shown along the abscissa and the number of listeners using different labels is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies splitting of the first tone, '2' implies splitting of second tone, 'N' implies perceived fusion , 'B' implies that both tones split.


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
PERCENT TRIALS
FREQUENCY DIFFERENCE (Hz) RE: nX400 Hz


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
-128 -64 -32 -16 -8 -4 0 4 8 16 32 64 128
0 1 2 3 4 5 6 7
NUMBER OF LISTENERS (OUT OF 7)
FREQUENCY DIFFERENCE (Hz) RE: nX400 Hz

Another **contextual factor** that may have brought the 'B' judgments is the **"streaming"** of sequential tones that are proximal in frequency. The "standard" harmonic of the first tone, in some of the stimuli of the present experiment may have streamed with the mistuned component of the second tone, by virtue of their being closer in frequency than any other adjacent component. The "component" stream would then segregate perceptually from the stream formed by the coincident, unchanged harmonics of the two tones.

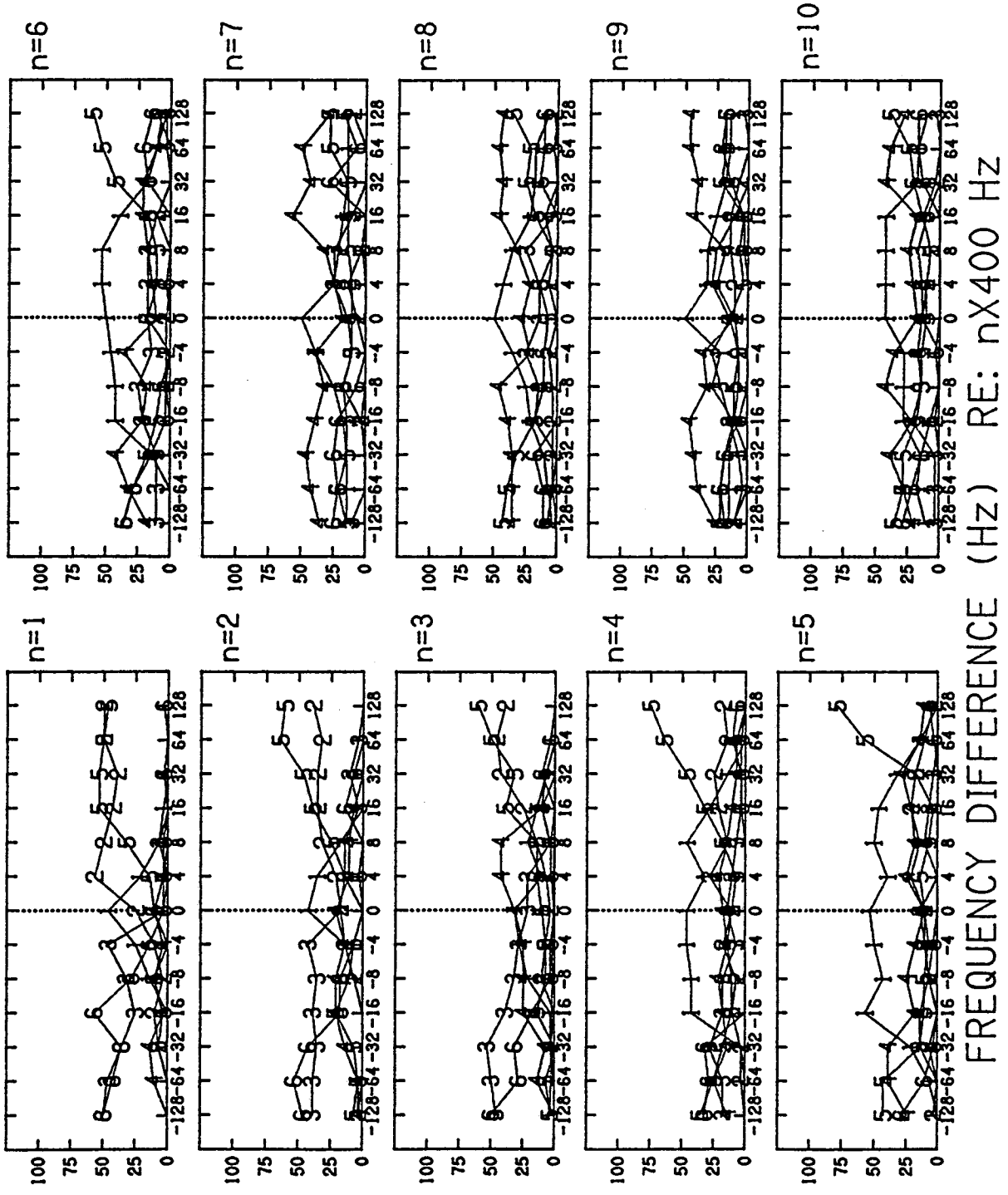
The report of listeners that a "melodic interval" was heard out against a low, dense background sound indicates that this type of perceptual grouping process may indeed have been at work. McAdams (1984 b) also reported that "melodies on partials" were sometimes heard, with stimuli of the type used here. Stream segregtion and other contextual factors are discussed in more detail in chapter 6.

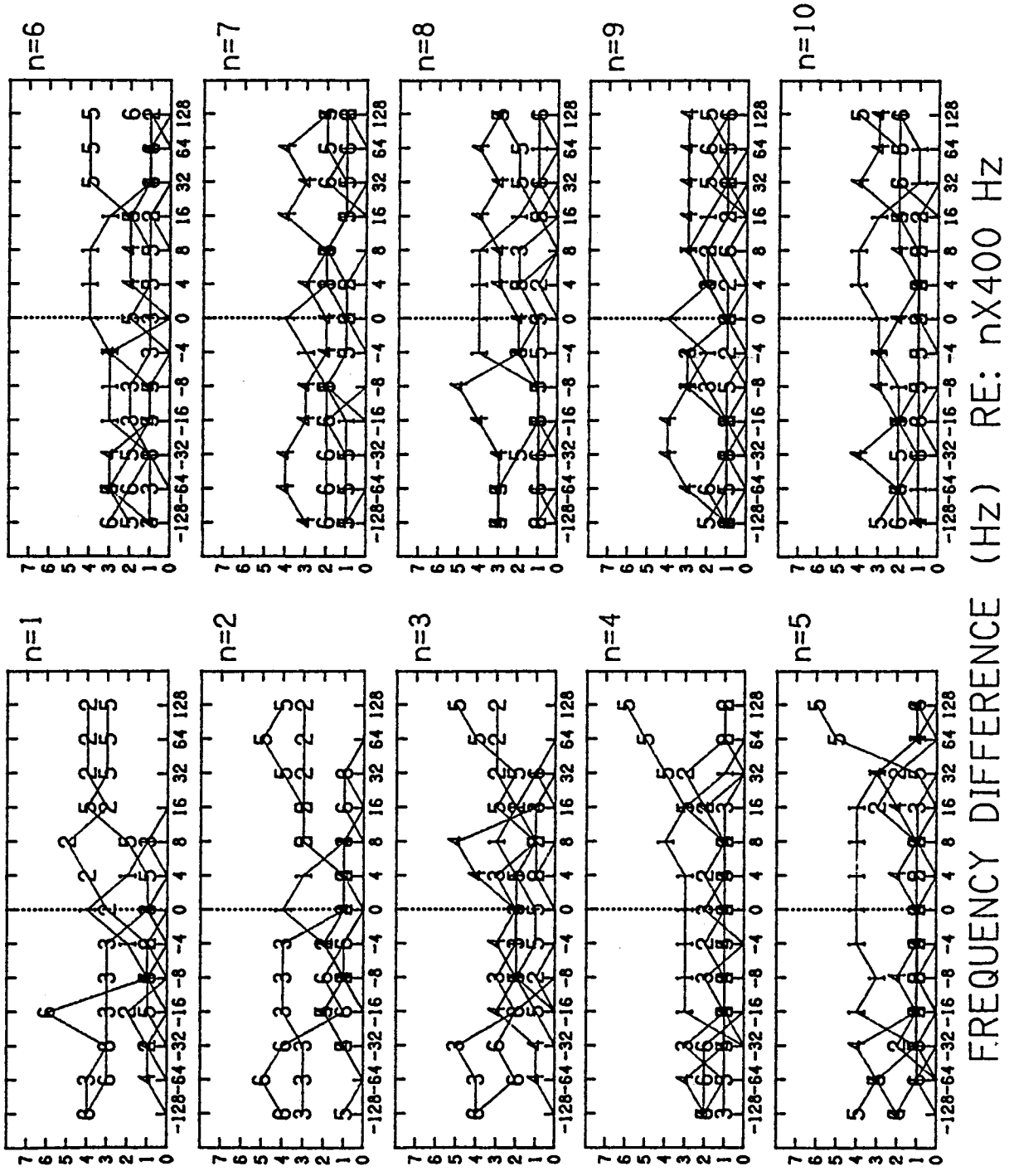
5.8.2 Second judgment responses

For the second judgment (figures 5.14 and 5.15), the main difference between the 200 and 400-Hz conditions is the greater use of label '4' for higher harmonics ($\geq$n=6) in the latter case. For the 200-Hz condition (figures 5.5 and 5.6), changes in frequency of lower harmonics ($<$ n=6) were correlated with changes in pitch and "something else", but higher harmonics did not evoke any single representative response label.

Figure 5.14 (next page) Data showing label choice (averaged over 7 listeners) for the **second judgment** for sequences with **linear** changes in the frequency of components of magnitude as shown along the abscissa (**re: nX400 Hz**). Proportion of trials on which the labels were used is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies no change, label '2' implies a perceived rise in pitch, label '3', a fall in pitch, label '4', a change in "something else", label '5', a rise in pitch and a change in "something else", label '6', a fall in pitch and a change in "something else".

Figure 5.15 (page 337) Distribution of dominant response labels across listeners for the **second judgment** for stimuli with **linear** changes in a single component **re:nX400 Hz.** Magnitude of change is shown along the abscissa and the number of listeners using different labels is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies no change, label '2' implies a perceived rise in pitch, label '3', a fall in pitch, label '4', a change in "something else", label '5', a rise in pitch and a change in "something else", label '6', a fall in pitch and a change in "something else".


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
PERCENT TRIALS
FREQUENCY DIFFERENCE (Hz) RE: nX400 Hz


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
-128 -64 -32 -16 -8 -4 0 4 8 16 32 64 128
0 1 2 3 4 5 6 7
FREQUENCY DIFFERENCE (Hz) RE: nX400 Hz
NUMBER OF LISTENERS (OUT OF 7)

For the 400-Hz condition, the label '4' emerges as the "dominant" response, indicating that a change in "something else" was perceived, unaccompanied by a change in timbre. This label may be taken to imply timbre change in this case, as "pitch" and "splitting" judgments were not a confounding factor for these higher harmonics (n=7 to 10).

5.9 Results for ratio changes with standard F0=400 Hz

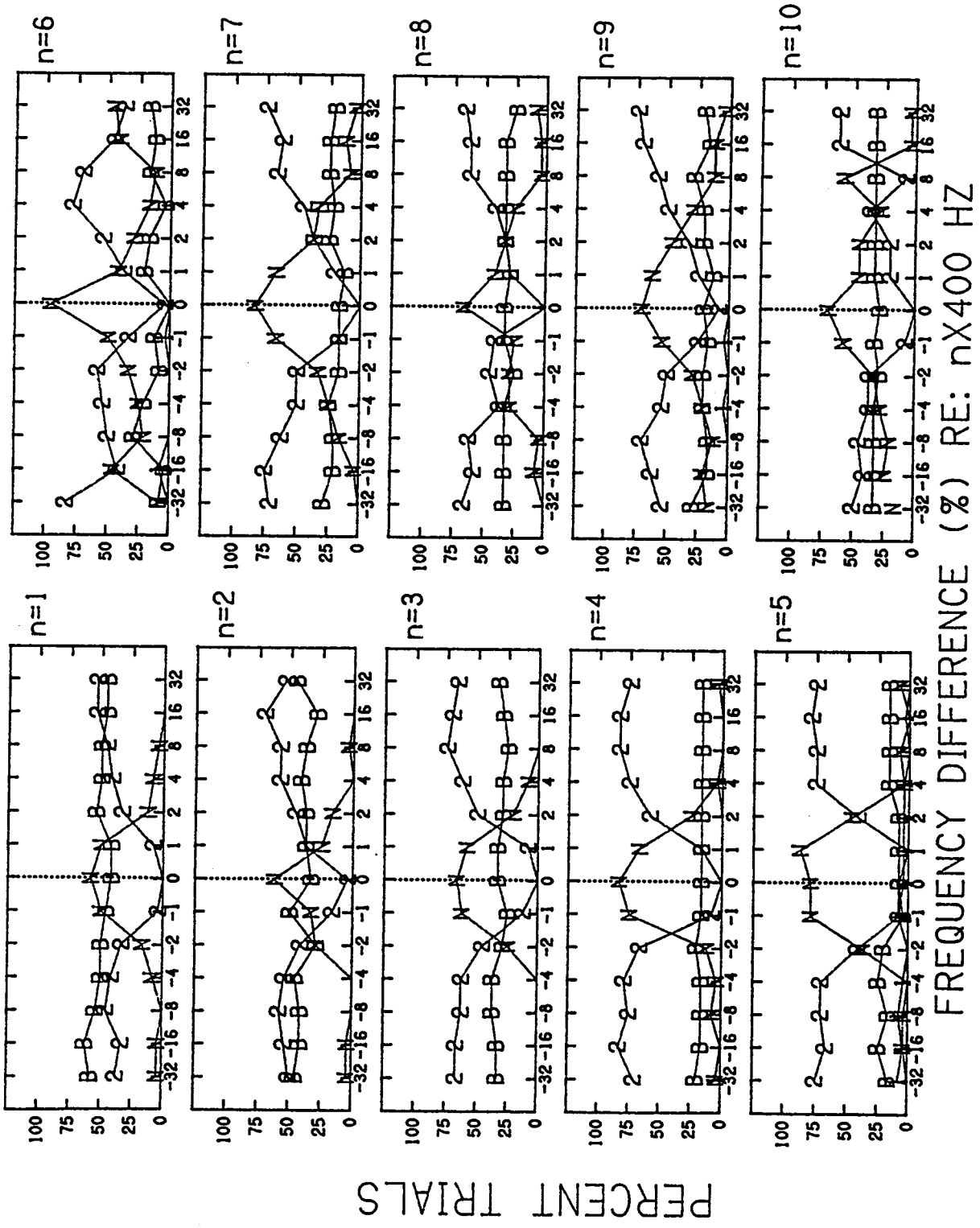
Figures 5.16 and 5.17 show the labelling data for the first judgment with ratio changes (in terms of % trials and number of listeners, respectively). Figures 5.18 and 5.19 show similar displays for the second judgment.

The pattern of results for the first judgment with F0=400 Hz is very similar to that obtained for F0=200 Hz. Segregation was observed over a greater range of harmonics as compared to the linear condition, and analytic listening was similarly enhanced, with 'B' being used on 25%-50% of the trials. The pattern of results for the second judgment was also similar to the 200-Hz condition (figures 5.10, 5.11), with labels '5' and '6' being used to imply changes in pitch and "something else" for $\Delta f \geq 4\%$, and considerable variability observed for smaller deviations.

In contrast to the second judgment for linear changes (figures 5.14, 5.15), the label '4' did not dominate for all ratio changes in $n \geq 6$. It was used predominantly for small values of ratio change ($\pm 1\%$).

Figure 5.16 (next page) Data showing label choice (averaged over 7 listeners) for the **first judgment** for sequences with **ratio** changes in the frequency of components of magnitude as shown along the abscissa (**re: nX400 Hz**). Proportion of trials on which the labels were used is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies splitting of the first tone, '2' implies splitting of second tone, 'N' implies perceived fusion , 'B' implies that both tones split.

Figure 5.17 (page 341) Distribution of dominant response labels across listeners for the **first judgment** for stimuli with **ratio** changes in a single component **re:nX400 Hz.** Magnitude of change is shown along the abscissa and the number of listeners using different labels is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies splitting of the first tone, '2' implies splitting of second tone, 'N' implies perceived fusion , 'B' implies that both tones split.


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
PERCENT TRIALS
FREQUENCY DIFFERENCE (%) RE: nX400 HZ

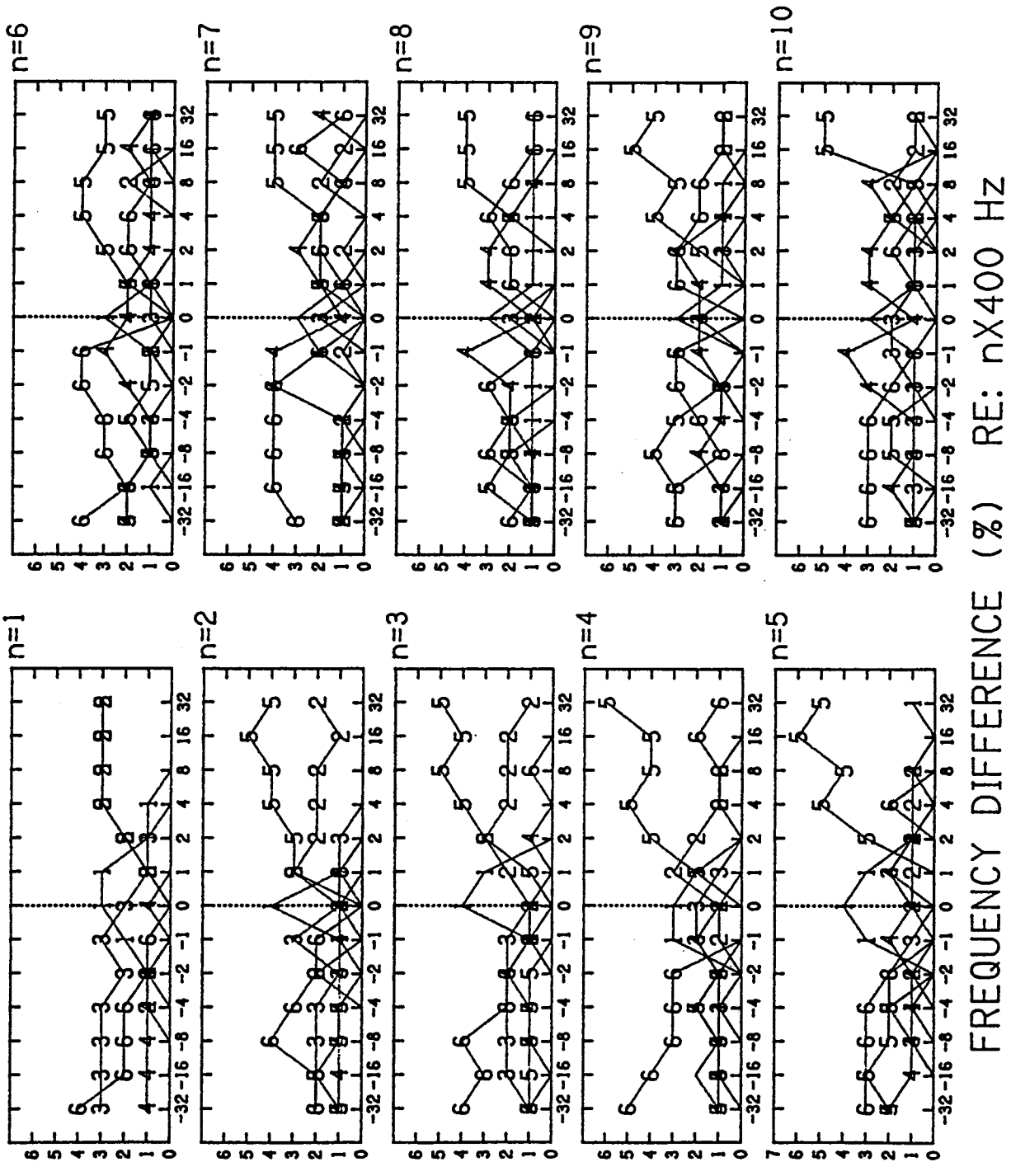
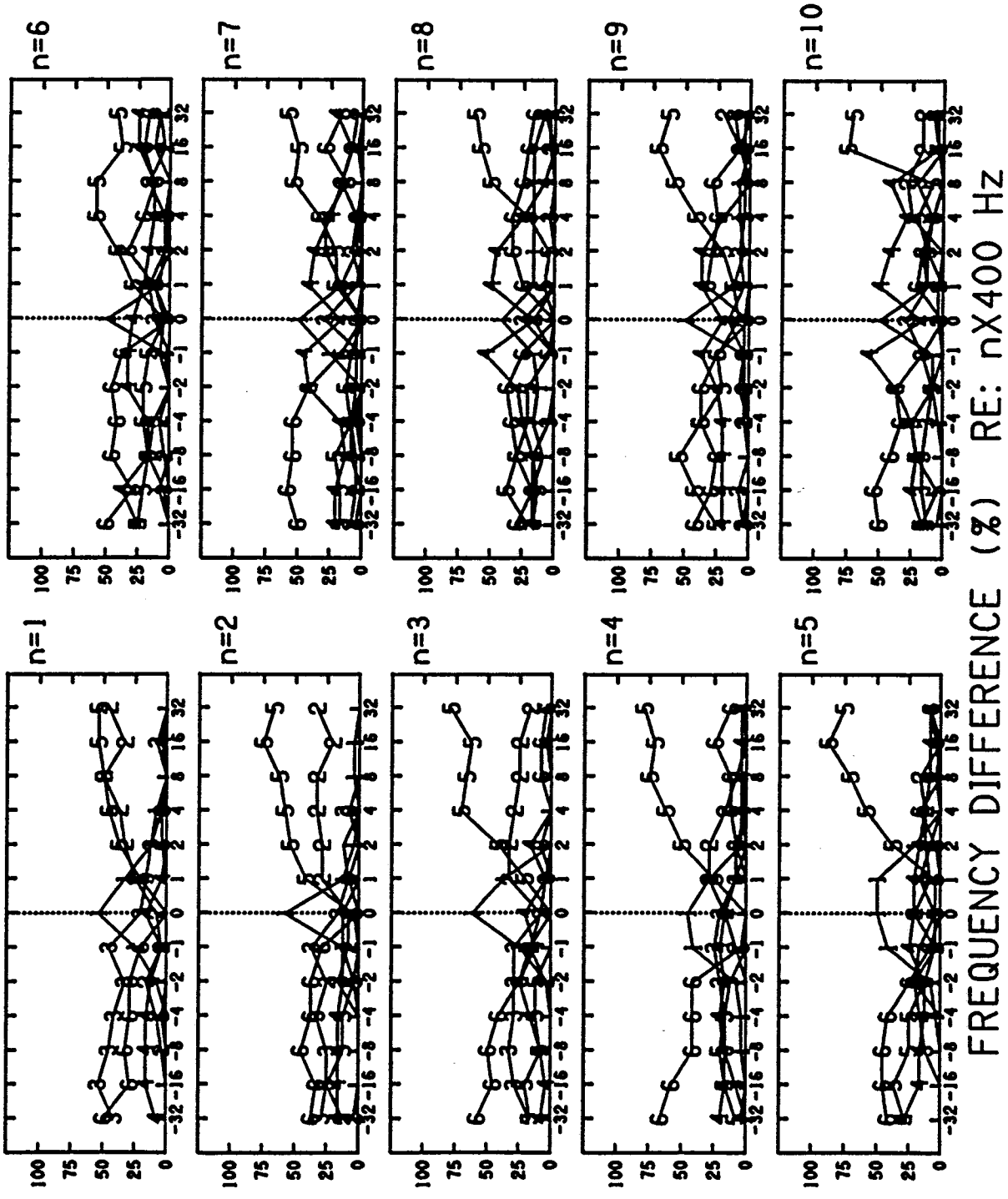
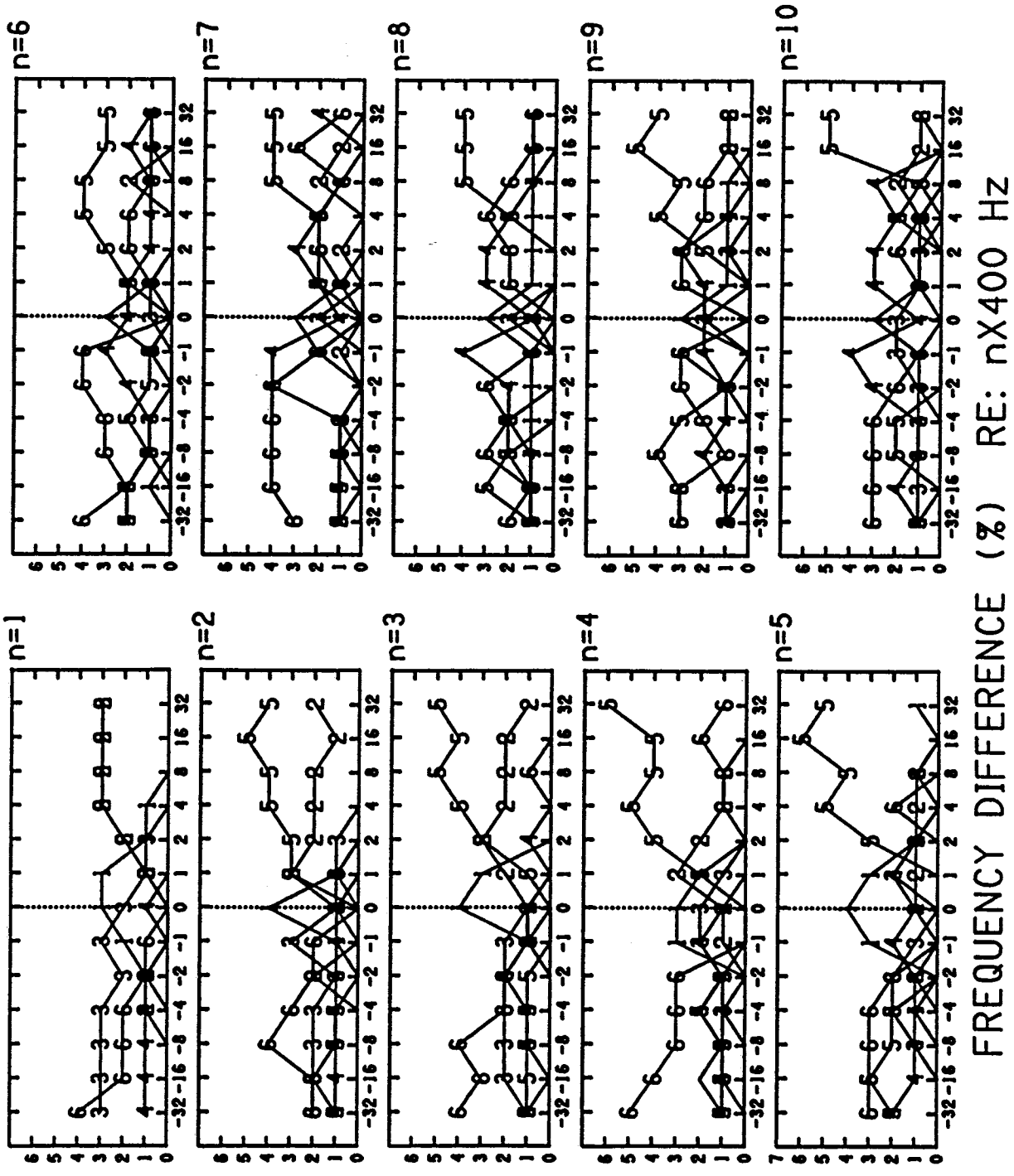

n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
NUMBER OF LISTENERS (OUT OF 6)
FREQUENCY DIFFERENCE (%) RE: nX400 Hz

Figure 5.18 (next page) Data showing label choice (averaged over 7 listeners) for the **second judgment** for sequences with **ratio** changes in the frequency of components of magnitude as shown along the abscissa (**re: nX400 Hz**). Proportion of trials on which the labels were used is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies no change, label '2' implies a perceived rise in pitch, label '3', a fall in pitch, label '4', a change in "something else", label '5', a rise in pitch and a change in "something else", label '6', a fall in pitch and a change in "something else".

Figure 5.19 (page 344) Distribution of dominant response labels across listeners for the **second judgment** for stimuli with **ratio** changes in a single component **re:nX400 Hz.** Magnitude of change is shown along the abscissa and the number of listeners using different labels is shown along the ordinate. Each frame corresponds to the number 'n' of the harmonic being displaced. Label '1' implies no change, label '2' implies a perceived rise in pitch, label '3', a fall in pitch, label '4', a change in "something else", label '5', a rise in pitch and a change in "something else", label '6', a fall in pitch and a change in "something else".


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
PERCENT TRIALS
FREQUENCY DIFFERENCE (%) RE: nX400 Hz


n=1
n=2
n=3
n=4
n=5
n=6
n=7
n=8
n=9
n=10
-32 -16 -8 -4 -2 -1 0 1 2 4 8 16 32
FREQUENCY DIFFERENCE (%) RE: nX400 Hz
NUMBER OF LISTENERS (out of 6)

5.10 Discussion

The results of the present experiment, (like those of experiment 2), validate the multiplicity of percepts associated with spectral changes in complex tones. Stimuli of the type used here have been studied by other invesigators as well (Hartmann, 1988; McAdams, 1984 a,b; Moore et al, 1984, 1985 a,b, 1986). Some of their experiments were reviewed in chapter 2 (section 2.10). It was mentioned that the threshold values of frequency change required for discrimination or identification judgments, and the variation of threshold with harmonic number showed different functional relations, depending on the task assigned.

In the present experiment, the double-layered task enabled simultaneous reports of changes in fusion, pitch and timbre. The relation between magnitude of change, harmonic number, and the perceptual change evoked could thus be studied directly. Thus, for example, it could be observed that:

i). Changes in the frequency of lower harmonics are more likely to yield splitting than changes (of equal physical magnitude) in higher harmonics.

ii). For ratio changes (of greater physical magnitude for higher components), splitting is observed across a wider range of harmonic numbers.

iii). Changes in the frequency of lower components elicit

judgments of pitch change.

iv). Changes in the frequency of higher components are more likely to elicit judgments of timbre change (for linear shifts).

The most salient perceptual change associated with the stimuli of the present experiment was the perceptual "fission" of a harmonic complex tone into more than one entity, when one harmonic was mistuned from its normal frequency. There was some listener variability in the present experiment for the second judgment pertaining to pitch and timbre change. For splitting judgments, however, most listeners (typically 71%) were in agreement with each other.

Factors promoting perceptual fusion of components of a complex sound were discussed in section 2.4.1. Harmonicity and correlation of frequency change across members of a spectral subgroup are two such factors that contribute to the fusion of the group into a unified "source image" (McAdams, 1984 b). The formation of a fused image simultaneously facilitates its perceptual separation from other spectral components that may be present in the auditory environment.

The perceptual "standing out" of components in the present experiments, may be an outcome of this type of a parsing process. The inharmonic complex formed by mistuning a component may split up perceptually, into a harmonic complex with pitch equivalent to F0, and the mistuned component with its own pitch, given adequate magnitude of mistuning.

However, the magnitude of change is not the only factor operative in eliciting judgments of splitting: Rather, the changed location of the displaced component within the framework of the other unchanged harmonics, influences splitting judgments. If the changed component moves beyond the bounds of the inter-component spacing, it may not yield a splitting judgment as it may be within the sub-threshold range for the nearest component.

The second task used in the present experiment proved to be limited, however, in that it did *not always* indicate clearly, if the pitch change reported referred to the pitch of the segregated component, or to that of the complex as a whole. In some cases however, it was clear that the pitch of the mistuned *component* was being "tracked", rather than that of the complex.

This was revealed when ***divergent directions of pitch change*** were reported. An upward deviation in frequency was sometimes reported as a fall in pitch, along with a change in "something else" (label '6'), and vice versa. Close scrutiny of such stimuli revealed that comparisons were most likely being made across the *components* of adjacent tones. Thus, if the mistuned harmonic 'n' of the second tone was displaced to a frequency value closer to that of harmonic 'n+1' or 'n-1' of the first tone, the pitch change often corresponded to the direction of frequency change between the sequential components that were nearer in frequency. The "wrong" harmonics 'n', and 'n+1' thus appear to have been compared (instead of 'n' and 'n') for these divergent judgments.

Frequency proximity of components of adjacent tones also appeared to play a part in judgments of both tones splitting (label 'B'). It is suggested that a **grouping principle** such as "stream segregation" led components close in frequency to "cohere" together in a perceptual s tream (Dannenbring and Bregman, 1978; van Noorden, 1975).

While magnitude and location of frequency changes have been studied by other investigators using complex stimuli of the type used here, context effects of the type discovered in the present experiment, have not been mentioned in most reports.

5.10.1 Comparison of results with those of McAdams (1984 b) and Moore et al. (1984, 1985 a, b, 1986)

The experiment by McAdams (1984 b) mentioned in the introduction (section 5.1) investigated the "effects of frequency modulation incoherence, harmonicity and intensity on multiple source perception". Stimuli of the type used in the present experiment were used, except that the change in frequency of a component was dynamically modulated. Also, the background components against which the shift was made, were either harmonically or inharmonically related. "Source multiplicity thresholds" were found to decrease with harmonic number.

If "source multiplicity" is construed as the perception of more

than one entity (as in the split judgments of the present experiment), this result is divergent from that obtained in the present experiment. For linear changes, the higher harmonics were more likely to remain fused, while the lower components segregated. For ratio changes, the splitting function was fairly flat across harmonic numbers.

From the discussion of "perceptual effects" related to deviations in different partials (McAdams, 1984 b, p. 106), it appears that percepts such as "roughness" and a "chorus" effect were construed as indicating more sources than the "emergence of a pitched sinusoid". Given this new interpretation to "source multiplicity", the results of the present experiment would agree with those of McAdams, in that the use of label '4' indicating timbre changes (such as "roughness") increased for the higher harmonic numbers.

The difference in our results and those of McAdams highlights the fact that different perceptual cues may underlie discrimination judgments for stimulus changes made along a single dimension. The extent to which different perceptual cues are actually employed in making a discrimination judgment depends to a large extent on the task assigned to the listener, and the comprehension of the task by the listener.

Several experiments by Moore et al. with stimuli of this type, reviewed at various points in this dissertation (e.g. in section 2.10) also indicate that thresholds of frequency change or mistuning required for discrimination or identification judgments, and the variation of threshold

with harmonic number show different functional relations, depending on the task assigned (see figure 2.11) and factors related to auditory frequency analysis, spectral fusion, and the processes involved in derivation of pitch and timbre.

Figure 2.11 showed that frequency DLs were higher for high harmonics when the task entailed reporting a pitch change (Moore et al., 1984). The frequency deviation thresholds were also found to increase with harmonic number for hearing mistuned partials as separate components (Moore et al., 1986). For detection of inharmonicity, however, the figure showed a decrease in thresholds with increasing harmonic number (Moore et al., 1985 b).

Lower, better-resolved components are more susceptible to being "heard out" (von Helmholtz, 1877/1954; Plomp and Mimpen, 1967). They also play a more dominant role in the pitch derivation process (Goldstein ,1973; Ritsma, 1967). Poorly-resolved high components, on the other hand, are harder to hear out, and play a less dominant role in the pitch estimation process.

It is not surprising therefore, that Moore et al.'s listeners discriminated the higher harmonics poorly for judgments of pitch change and "hearing out". The lower thresholds for inharmonicity with increasing harmonic number appear to be related to the availability of timbral cues such as changes in "roughness" arising from mutual interference of unresolved components within critical bands.

The results of the experiments reaffirm the observations of Moore

et al. (1985 b), that there are discernible differences in the way in which changes in frequency of low and high spectral components are perceived.

5.10.2 Perception of phase changes

Another reason for the differences in the perceptual changes associated with high and low components can be found by looking at the stimuli from a temporal viewpoint.

The stimuli for experiments 2 and 3 were designed to differ in terms of spectral changes in the frequency of components. The temporal consequences of this manipulation were not a focus of stimulus design. In reality however, every spectral change has a consequence in the time domain. This is particularly important for the inharmonic stimuli of these experiments. As mentioned in section 4.6, the change in frequency of one or more components of a complex leads to a running phase change that causes the fine structure of the waveform to drift within the lobes of the overall temporal envelope. These phase changes are manifested as dynamic modulations of the waveform (see **figures 4.11 and 5.20**).

In figure 5.20, two pairs of waveforms are shown. The waveform for the standard, harmonic tone (F0=200 Hz) is shown on the left in both cases (time not to scale). The waveform on the top right is for a 10-component complex with the second harmonic mistuned in frequency by +32 Hz. The waveform at bottom-left is for a 10-component complex with the sixth harmonic mistuned by +16 Hz.

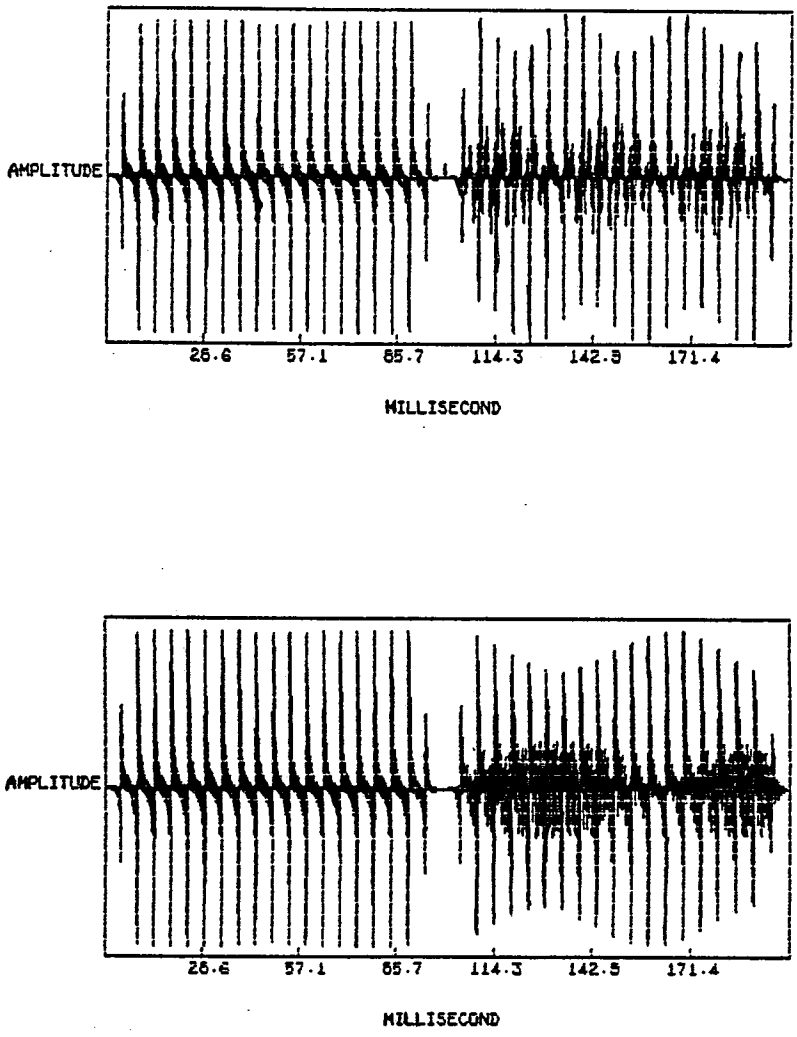


Figure 5.20 Waveforms for two sequences used in experiment 3 (sharpness of the peaks is an artifact of the digital scaling program used). The two sounds of a stimulus pair are shown side by side for comparison (not to scale in time). The waveform on the left is for the "standard" harmonic complex tone with F0=200 Hz in both pairs. The waveform on the top-right is for a 10-component complex with n=2 shifted by 32 Hz re:nX200 Hz. The waveform at bottom-right is for a 10-component complex with n=6 shifted by 16 Hz re:nX200 Hz. The harmonic sound exhibits regular periodicity, while modulation of the inharmonic sound is clearly visible.

As discussed in the section 2.7.4 on "auditory spectral filtering and phase perception" the relative amplitude fluctuations of the waveform can be perceived as roughness. This type of timbral cue would not be available for the well-resolved low components. Lower thresholds for inharmonicity with increasing harmonic number may thus be related to such changes in timbre.

5.11 Conclusions:

The myriad ways in which spectral changes can be made, and the range of percepts that accompany them often elude empirical investigation. The provision of the double task in our experiments, and the options available in each allowed the issue of multiple roles of a mistuned harmonic to be explored. Changes in perceived **fusion, pitch and timbre** were observed.

Changes in pitch appeared to be related to the magnitude of shift when made in the frequency of low components ($\leq n=6$). Changes in timbre appeared to be related to lack of resolution of high harmonics. Lack of fusion appeared to be a direct consequence of the inharmonicity of the stimulus.

However, factors other than the *mere induction* of inharmonicity also appeared to play a part in the parsing of a complex sound. Some of these factors are:

1). Magnitude of frequency change

2). Location of frequency change

3). Spectral context - both *within* and *across* the complex sounds

While the first two factors have been the focus of other studies with stimuli of the type used in the present experiment, the third factor was revealed in the present experiment, via the two magnitudes of frequency change, and the multiple options provided in the double-layered task assigned to listeners.

The larger-magnitude ratio changes made in the frequency of some high harmonics (most clearly for n=6), revealed that frequency proximity of the displaced component relative to the unchanged harmonic components influenced splitting judgments. Increase in the deviation of a component resulted in more judgments of splitting until the displaced component began to approach another "fixed" component, at which point it was no longer perceived as being "split". The spectral context within the complex was thus an important factor in segregation.

Both sequential and simultaneous principles of organization may come into play in derivation of the perceptual changes associated with frequency changes in harmonics within complex tones. The results of this experiment and the former two, are reviewed in the next chapter, within this larger perspective of perceptual organization.

Chapter 6
RECAPITULATION

6.1 Thematic review

Perceptual organization of complex sounds is a basic accomplishment in music and speech perception. The acoustic environment in which we live, typically comprises the combined contributions of various sound sources present simultaneously and sequentially. Despite this physical complexity, we are able to sort out the acoustic input into coherent perceptual units and retrieve the underlying structure of the contributions of individual, primary sources.

This marvellous sorting of a motley acoustic input into perceptual structures such as words, notes, chords, phrases, melodies, and rhythms reflects the impressive skill of the auditory system to analyze, synthesize, seek and find relations between acoustic "events".

Such comparisons may be made along primary physical dimensions of sounds, such as frequency, amplitude, phase and duration, or higher order percepts such as overall pitch, timbre, loudness and perceived duration. While these percepts are usually evoked as properties representing a "group", they in turn can influence how groups should be formed. Thus, a group of partials may be fused into a single entity with a particular pitch and timbre, but relations between pitches and timbres may determine the perception of other sound groups such as melodies,

instrument families, etc.

The effects of competing differences in pitch and timbre on melodic "channeling" or "streaming" of a sequence of complex tones were studied in an earlier experiment on perceptual grouping of **sequences** (Singh, 1984/1987). The results indicated that features of sounds were being compared not just at the "note" level, but also at the *component* level within and across notes. **Hierarchical grouping strategies** appeared to be operating in the grouping of components into fused sounds (or "notes") and further, of notes into melodies.

Further, the perceived relations between the pitch and timbre of the complex tones appeared to be influenced by the context of spectral components in *neighboring* complexes. It was thus indicated that these percepts may not be steady, passive features of a sound event, but rather, are dependent on dynamic **relations** between the components present simultaneously within the complex as well as on relations (such as frequency proximity) with components of temporally-proximal tones.

The experiments reported in this dissertation were designed to probe the processes that guide a listener's ability to differentiate components of a compound acoustic input into perceptual groups and assign them features such as timbre and pitch. Complex sounds of different spectral design were used as stimuli and a task designed to ascertain the nature of perceptual changes associated with changes in the frequency of one or more spectral components.

The choice of working with a **pair** of sounds rather than a longer

sequence, reduced the range of possibilities for stimulus variation across time to the simple case of two sounds being selected to be same or different. This type of scenario is typical of **discrimination** experiments in which listeners are required to report if two sounds are same or different, or to identify **which** sound is different from a "standard" sound. The task employed in the present experiments differed from such simple discrimination tasks in that listeners were required to indicate not just **if** two sounds were different, but also to identify ways in which they differed.

The many experiments of Moore et al. (reviewed in Chapters 2 and 5), indicated that changes in the frequency of harmonics of complex tones may be construed as changes in pitch, changes in timbre related to "inharmonicity", changes in the number of perceived sources (with the displaced component acquiring a separate identity), or a combination of these percepts. The very fact that it took Moore et al. four different looks at the *same* basic stimuli to establish the range of percepts associated with changed frequency of components, highlights the point that **identifiable perceptual cues** can be used by listeners in making what on the surface appear to be generically opaque discrimination judgments.

The extent to which different perceptual cues are *actually* employed in making a discrimination judgment depends to a large extent on the task assigned to the listener, and the comprehension of the task by the listener. Thus, a conscientious listener assigned a same/different task or a 2I2AFC task in a **pitch** discrimination experiment may ignore

discernible differences between stimuli if they are not explicit changes in pitch. Conversely, a shrewd listener may discriminate stimuli based on the slightest timbral differences even in the absence of a perceptible pitch change. Such a listener may achieve a high performance score despite not having done the assigned task, while the conscientious listener may receive a poor score.

This was precisely the finding of an experiment by Lauter and Singh (1985). The task in their experiment asked for reports of pitch changes for complex tones of different spectral density. While pure tones and harmonic complex tones with F0 change probably did provide reliable pitch cues for discrimination, such cues were absent for changes in the frequency of a single component embedded in a dense complex. The perceptual correlate for the latter type of stimulus was more likely a timbral cue such as "buzziness". This was revealed when a listener who performed near-perfectly with pure tones and harmonic complexes appeared to fail in discriminating sounds in which a single higher-amplitude component was changed in frequency by Δf. Conversely, a listener who performed poorly on the pitch discrimination task with clear, tonal stimuli, suddenly improved in discriminating the more complex stimuli. The former stimuli provided unambiguous pitch cues unlike the latter. However, there were fairly noticeable changes in quality in the latter stimuli. The two listeners had obviously adopted different perceptual criteria to discriminate between the sounds --- pitch differences in the first case, and timbre change in the second.

Experimenters thus need to be cautious about the design of discrimination experiments. Providing feedback for a task based on strictly perceptual criteria (such as pitch change), may enable a listener to learn to discriminate changes that, ironically, do not employ that percept !

The difficult task of describing a perceptual change was approached in the present experiments by sidestepping open-ended verbal reports, avoiding feedback, and resorting instead to a set of options (illustrated visually and via acoustic examples) that seemed to cover the range of perceptual changes.

Changes made in the frequency of one or more components of a complex tone in a sequential context, possess the potential to yield perceptual changes encompassing differences in perceived **pitch, timbre and fusion** of the altered complex. In some cases, changes in more than one of these attributes may simultaneously be perceived. The experiments sought to **map out** this range of percepts and their relation to physical features of the stimuli.

6.2 Summary of experiments

Three types of complex sounds were used as stimuli in three related experiments:

(1) harmonic "residue" tones comprising 4 components,

(2) 10-component harmonic and inharmonic complexes with

all components shifted away from some reference frequency,

(3) 10-component complexes with a **single** "mistuned" component displaced from its harmonic frequency.

The sounds were presented in pairs, with the first always being a "standard", relative to which frequencies of components of the second sound were shifted. Labelling tasks were especially designed to explore the range of percepts associated with stimulus changes.

When harmonic residue tones were used, the results of the first experiment showed that spectral locus is generally correlated with timbre, while spectral spacing governs pitch. However, the analytic mode of listening may interfere with the perception of synthetic "residue" pitch. This was indicated by the influence of absolute frequencies of components on the perception of pitch change. In the absence of changes in spectral spacing, but given changes in spectral locus, many listeners reported a simultaneous change in both pitch and timbre despite the fact that the fundamental frequency was unchanged, or in some cases, changed in a direction *opposite* to that of the reported pitch change. An interaction of pitch and timbre is thus implied. Both **relative** and **absolute** aspects of the spectrum appear to be taken into account in the derivational processes that assign these percepts to groups of spectral components.

The role of relative and absolute frequency differences between

spectral components was studied further in the second experiment. For complexes in which **all** components were shifted in frequency, the results indicated that changes that preserve harmonicity are generally perceived as a rise or fall in the pitch of a fused complex tone. However, if the changes do not maintain harmonicity, the complex may be heard as a diffused collection of overlapping sounds with ambiguous pitches. Such changes in frequency may also be perceived as causing a change such as "roughness", in the timbral quality of the complex. Small linear deviations in frequency generally resulted in a fused percept with a slight change in pitch. Deviations beyond 4% of the spectral spacing, however, led to diffusion and perceived changes in timbre.

In light of differences in the absolute and differential frequency sensitivity of the auditory system as a function of frequency range, and in the contribution of different spectral regions to the perception of pitch, the perceptual correlates of changes in the frequency of **individual** spectral components were investigated in the third experiment. Only **one** component of the comparison sound was changed in frequency. The remaining components retained their harmonicity. The perceptual consequences for this type of change were different, depending on the location and magnitude of the change. Different magnitudes of frequency shift were used.

For **"linear" changes**, the displacement in frequency was always within the bounds of the spectral spacing of the components (i.e. a maximum of 32% of the spacing). For such changes made in the

frequency of components of low harmonic order (n=1 to 3), the mistuned component was heard as a separate tone with its own pitch , while the remaining components were heard as a fused complex tone. Such linear displacements, when made in the frequency of higher components, however, did not lead to segregation. Rather, they were more likely to yield a change in a timbral attribute such as roughness.

For **"ratio" changes**, different harmonics were compared for frequency shifts of the same "relative" magnitude proportional to the standard harmonic frequency. Different magnitudes of shift thus ensued in terms of "absolute" differences in Hertz across components. Since the ratio changes were generally greater in magnitude than the "linear" changes, splitting was observed over a greater range of harmonic numbers. However, it was observed that for some harmonics, these shifts led to changed frequency values that exceeded the spectral spacing (e.g. a change of $\pm$32% in the 4th harmonic of 200 Hz led to a shift of $\Delta f = \pm 256$ Hz that exceeded the 200 Hz spacing. The new component thus crossed the adjacent n=5 (i.e. 1000 Hz) and the n=3 (i.e. 600 Hz) boundary as its frequency was raised or lowered to 1056 Hz or 544 Hz, respectively).

The changed component was perceived in the overall **context** of the components within the complex. Different magnitudes of shift led to changed relationships between the new shifted component and the unshifted harmonics. In some cases, the mistuning resulted in a **split** percept, while in some cases the new component appeared within **subthreshold** values of the splitting threshold for another harmonic

number, and was thus not heard as segregated.

A further interesting indication of **contextual influence** was exhibited by some listeners report of perceived splitting in the first "standard" tone of a stimulus pair. While this effect was not dominant, in that it was reported only on 25 to 50% of the trials for some stimuli, it is nevertheless curious. The first tone of the stimulus pair was always a harmonic complex and remained unchanged across trials. Harmonicity usually facilitates fusion of components, as was discussed in chapter 2. However, listeners sometimes reported hearing a segregated component in the first sound event, followed by another one of different pitch in the second sound event. A segregated melodic fragment was thus heard against a backdrop of the unchanged components.

The unexpected segregation of the first complex implies a type of "retrospective" comparison of individual components within the two complexes. The target harmonic of the first complex appeared to form a perceptual group or **"stream"** with its changed counterpart in the second complex, while the unchanged harmonics formed a separate stream. Stream segregation, described in chapter 1, is discussed below. The main point to be noted from this particular outcome of experiment 3 is that analytic listening is dependent on both the content, and context of spectral components presented **simultaneously and sequentially.** A component that is normally camouflaged within a fused complex tone can acquire a separate perceptual identity, given the appropriate context. Perceived fusion or fission of sounds is thus affected not just by the

inter-component relations ***within*** a complex sound event , but by relations ***between*** temporally separated sounds events as well.

The provision of the double experimental task also enabled determination of timbre change and the direction of perceived pitch change (if any). Contextual listening was again indicated when listeners sometimes reported a pitch change in a direction *opposite* to that in which the frequency of a component was shifted. Instead of tracking the frequency of the changed harmonic 'n', it appeared that the "wrong" harmonic ('n+1', 'n-1' etc.) was compared for some magnitudes of change. Frequency proximity between individual components of the standard and changed complexes appeared to dictate **which** components were compared. Sequential context thus influenced the parsing of the simultaneous components.

The results of the second and third experiments verified that inharmonic complexes might not be perceived as fused entities with distinctive pitch and timbre attributes. Rather, the unitary sensation usually associated with a harmonic complex may be replaced by one of multiple sources present simultaneously. The range in which fusion or fission of components prevails was explored by systematically varying the magnitude, location, and type of spectral change (e.g. ratio or linear). Further, the labelling task also allowed investigation of pitch and timbre changes associated with the fusion or segregation of the complexes.

The results of the three experiments summarized above have already been discussed at length in the chapters describing them. In the

following sections, the implications of the findings are reviewed in **broader perspective** related to issues of pitch and timbre processing, fusion and segregation of sounds, general principles of perceptual organization, and the assignment of perceptual attributes to groups of spectral components.

6.3 The separability of pitch and timbre

Extant research on timbre and pitch was reviewed in considerable detail in chapter 2. It was noted that these percepts have been theoretically treated as being independent, despite observations that they influence each other. Pitch perception is fundamentally dependent on **"relative"** comparison of frequencies of spectral components while timbre is strongly associated with the **"absolute"** location of the spectrum. It seems however, that these percepts may be influenced to some extent by both types of comparison.

Timbre may be influenced by spectral spacing in addition to spectral locus. At low fundamental frequencies, spectral spacing is narrow. The high components of complex sounds with a low F0 may thus fall into the same critical band, leading to interference phenomena such as roughness and dissonance that are reported as timbre changes. The wider spacing of spectral components at higher fundamental frequencies, on the other hand, may lead to reduced fusion with certain spectral components becoming more audible than others. This changed weighting

of components may also be construed as a timbre change.

Similarly, pitch may be influenced by spectral locus, as well as by spectral spacing. Evidence already exists in the pitch shift data from the experiments of de Boer (1956/1976) and Schouten et al. (1962). In those experiments, pitch was seen to be influenced by both the location of spectral components, and their spacing. The experiments reported here also demonstrate this effect. In experiment 1, many listeners reported pitch changes when the spacing (or F0) of the complexes remained *unchanged.* The pitch change was correlated with a shift in the absolute position of the spectrum. In the inharmonic stimuli of the second experiment, spacing and general location of spectral components remained the same, while the F0 relation between them was disrupted. Again, pitch was influenced, independently of timbre for a small range of frequency change ($\Delta f \leq 4\%$) and coupled with timbre changes for larger deviations.

Since the frequency spectrum contributes to both the pitch and the timbre of a sound, spectral changes may be construed as changes in either pitch, or timbre, or both of these percepts simultaneously. The relation between **spectral pitch and timbre** was discussed in chapter 2 (section 2.9.3). It was suggested that the spectral pitch associated with the frequency of spectral components physically present in a complex sound contributes to both, the overall virtual pitch of a complex sound, as well as to its timbre. While the relation between spectral and virtual pitch was well established in the very genesis of these terms (Terhardt,

1974), the relation between spectral pitch and timbre has been articulated less strongly.

Experiments in which the spectral design of the stimuli or the experimental task involved comparison of sounds with **divergent spectral and virtual pitch information**, demonstrated the perceptual confusion between pitch and timbre (Davis et al., 1951; Plomp, 1967; Risset, 1971; Ritsma, 1966; Schouten et al., 1962; Smoorenburg, 1972). Since pitch was the percept of interest in these studies, the timbral consequences of changes in spectral frequencies were not reported rigorously.

The varied repertoire of responses permitted in the present experiments allowed percepts other than pitch to be reported as well. The use of these extended response options showed that listeners differ in their ability to separate timbre from pitch. Differences in spectral frequencies (or spectral pitch) are often equated with differences in both the overall virtual pitch as well as the timbre of the sound.

Risset (1971) and Smoorenburg (1970) also noted this type of diversity in the pitch judgments of their listeners. Different listeners were consistently able to follow changes in either the spectral pitch, or the "tonal" (virtual) pitch related to the F0. While Smoorenburg regarded the timbre difference as an interfering factor responsible for the confusion, Risset concluded that "the concept of pitch may not be the same for everybody".

The results of the present work taken in the context of earlier

studies further strengthen the idea that the percept we call "pitch" is dichotomous. Furthermore, it is indicated that the dimension of pitch that depends on absolute frequency is closely related to timbre. This appears to be the **"body"** pitch discussed by Davis et al. (1951), the **"place"** pitch described by Licklider (1954), and the **"spectral" pitch** described by Terhardt (1974). A similar dichotomy was also proposed for the perception of timbre by Schaeffer (1966). He used the terms "matiere" (material) and "forme" (shape) to distinguish between spectral and temporal factors in timbre perception. **"Material"** appears to be directly related to spectral or place pitch.

A change in spectral frequency or "material" would thus affect both the pitch and the timbre of a sound. This relation between spectral pitch, overall pitch, and timbre has not been acknowledged widely at present but is gradually gaining acceptance (Hirsh,1988; Risset 1978; Singh, 1987). Houtsma (1979, p. 98) suggested that the "constant rivalry between analytic and synthetic" modes of perception is responsible for the duality in the perception of pitch of complex tones. The results of the present experiments indicate that this statement needs to be extended to include the idea of rivalry between pitch and timbre as well.

6.4 Fusion and fission

Factors enabling the perceptual fusion of spectral components to form a unified complex sound were discussed in chapter 2. Harmonicity,

temporal synchrony, coherence of frequency and amplitude modulation are some of the physical criteria that have been identified by McAdams (1984 a,b) as facilitating such perceptual groups to be formed. The fusion of a group of related components into an "image" simultaneously facilitates its perceptual separation from irrelevant acoustic material that does not exhibit similarly-correlated behavior. The segregation of some of the stimuli used in experiments 2 and 3 seems to be a direct consequence of defiance of the principles that promote fusion.

The spectral changes that were focussed on in experiment 2 involved changing the frequency of **all** components, either maintaining the same ratio of harmonic frequencies relative to the F0, or maintaining constant differences between components. The latter type of change led to the complex becoming inharmonic. Such inharmonic stimuli were perceived as being diffused in accord with similar observations made by McAdams (1984 b).

In contrast to the second experiment, the stimuli of experiment 3 involved mistuning **one** component of a complex, while the others remained harmonic. The mistuned harmonic was clearly audible for these stimuli when there were large deviations in frequency. Such stimuli have been investigated by other researchers as well (McAdams, 1984 a, b; Moore et al., 1984, 1985 a, b, 1986; Hartmann, 1988). Comparative details of their experiments were provided in chapters 2 and 5. Possible **"mechanisms"** responsible for bringing about the perceptual splitting or fission of the sounds are discussed below.

The stimuli of experiments 2 and 3 provided two types of change in frequency of components: linear and ratio. While these types of changes differed only in magnitude of deviation when the frequency of a single harmonic was changed (as in experiment 3), they considerably altered inter-component relations when made *en masse* in the frequency of all components (as in the stimuli of experiment 2).

In the case of a **single** harmonic being mistuned, the frequency behavior of the other components still remains correlated in that they maintain their frequency ratios relative to the original F0. As suggested by McAdams (1984 b), the parsing of such a stimulus may have been brought about via a **"cross channel" mechanism** that looks for correlations in the output of different frequency-specific "channels". The detection of two uncorrelated periodicities may then be interpreted as being indicative of two different sources. Some physiological evidence for such periodicity detection has been provided by recordings of responses in auditory nerve fibers of cats stimulated with complex sounds (Evans, 1978).

When **all** components are changed in frequency by the same ratio relative to the individual harmonic frequencies, the harmonicity of the complex is preserved. Preservation of harmonicity implies the preservation of regularity in the temporal discharge pattern of responding auditory nerve fibers. A cross-channel coherence detection mechanism would thus infer that a single source was present, given the absence of differences across frequency channels. Such a mechanism

may have been responsible for the perceptual fusion of the harmonic stimuli used in experiment 2. However, even the harmonic stimuli were sometimes reported as being "split" at the largest magnitudes of frequency change (64% of the F0). In those cases, the fusion cue provided by harmonicity may have been counteracted by fission cues provided by the **flat** spectral envelope (atypical of most natural tonal sounds) and the wider spacing of spectral components (that promotes analytic listening (Plomp, 1976)).

For the inharmonic stimuli of experiment 2, the frequency of all components was changed by the same absolute amount but by different relative amounts. Such changes may be construed as being "correlated" in the sense that they were of equal physical magnitude, and made in the same direction ($\pm$). A cross-channel coherence detector should then have interpreted the changed sound as being a unified source. However, both our listeners, and those of McAdams (1984 b) perceived such stimuli as being segregated.

Frequency ratios have repeatedly been shown to be important in hearing (Viemeister and Fantini, 1987). It seems that they are also important in promoting fusion of components of a tonal complex. Failure to maintain harmonic ratios led to segregation despite the common direction of frequency changes in the components.

The proposed cross channel detector thus also appears to be sensitive to ratios amongst detected periodicities. Running phase changes ensuing from the lack of a common periodicity amongst inharmonic

components lead to irregularities in the temporal discharge pattern that may be interpreted as being indicative of multiple sources.

Temporal variations in the **physical** stimulus are also caused by interaction of spectral components close in frequency. Some of the frequency changes made in the stimuli of experiments 2 and 3 led to such variations as was shown in the modulated waveforms in figure 5. The interaction of components lying within critical bands would also lead to temporally-variant *neural* discharge patterns. Based on earlier suggestions by Goldstein (1966) and Schubert and Nixon (1970), McAdams (1984 b) has also proposed a **"within channel" mechanism** for the extraction of source information based on changes in the temporal discharge pattern *within* an auditory channel.

Many experiments have been conducted that show that information about the periodic structure of the waveforms of both pure and complex tone stimuli is preserved in the temporal discharge patterns of cochlear fibers (Rose et al.,1967; Hind et al.,1967; Brugge et al. 1969). McAdams suggests that irregularities or perturbations of periodicity in such patterns may be used to signal the presence of multiple sources.

The two gradations of frequency change (linear and ratio) provided in the stimuli of experiments 2 and 3 may have provided such within-channel cues for some magnitudes of deviation that led to components interfering within critical bands. For the stimuli in which **all** components were changed, it is not possible to ferret out which reports of "splitting" were based on cross-channel and which on within-channel

mechanisms. For stimuli in which single components were shifted, however, it seems reasonable to assume that splitting responses were based on a cross-channel mechanism that resulted in grouping of harmonics into one fused sound, perceptually separable from the deviant component.

6.5 Fusion, pitch and timbre

The type of within-channel interaction alluded to by McAdams may also play a role in conveying timbre changes. Variations in the waveform of inharmonic stimuli have been associated with changes in timbre construed as roughness (Plomp, 1970; Terhardt, 1974; Patterson, 1972). Within-channel perturbations cued by irregularities in temporal discharge patterns could thus be construed as changes in timbre.

The double task in our experiments allowed reporting of timbre changes. Linear changes made in the frequency of higher components were indeed correlated with timbre change as may be expected, given the greater spectral density and likelihood of interference within critical bands.

The double task further showed that small deviations in frequency in experiment 2, were construed as changes in pitch, while larger deviations were correlated with both pitch and timbre changes. In keeping with modern theories of pitch perception (Goldstein, 1973), Cohen (1980) noted that: "If a stimulus cannot be made to fit a standard

template, individual pitches will emerge rather than a unitary pitch". The perception of multiple "sources" in the present experiments seems to have been subserved by such an emergence of multiple pitches.

Cohen (1980) also addressed the question "whether tones perceived as dispersed can have any tonal meaning" ? While the task employed in our experiments did not call for elaborate musical judgments, the detection of a pitch change for the dispersed stimuli implies that they too could potentially be put to musical use. However, the dispersed nature of such stimuli would vary as frequency relations were changed. Additional changes in timbre and pitch may then ensue. Also, such stimuli would not be amenable to description in terms of a common frequency related to a global pitch. Musical composition with such inharmonic stimuli would thus necessitate construction of a new referential system with which to scale differences between stimuli. Such attempts have been made by Slaymaker (1970) and Mathews and Pierce (1980) with some measure of success.

6.6 The influence of context and stream segregation on audibility of components

The splitting data obtained for experiment 3 are in accord with the **"Newtonian law of fusion"** coined by Cohen (1980): "a harmonic tone will remain fused unless some outside force like a sudden change in

one partial, or some inner mechanism such as memory for the sound of a partial explicitly directs attention to a portion of the total structure" (p. 27-28) !

An "outside force" was indeed applied in the stimuli of experiment 3 in terms of the mistuning of one harmonic and led to segregation. The latter idea of directed attention has been mentioned by von Helmholtz (1877/1954) and was applied experimentally by Plomp (1964). Plomp provided a comparison tone equal to the frequency of a component to be used as a reference in "hearing out" the component when it was embedded in a tone complex. The sequential context of this 'pointing' tone seems to have been helpful in the hearing out of the target harmonic. The contextual observations described in section 6.2 also indicate a beneficial contribution of pertinent sequential information in organization of sounds. It was suggested that "stream segregation" played a part in this type of sequential information processing.

Streaming occurs when sounds differ ing from each other along some dimension are presented in rapid succession. The original sequence is often perceptually segregated into subsequences or "streams" within which the range of differences is smaller (Bregman and Campbell, 1971). Stream segregation was described briefly in chapter 1 along with a review of the author's previous work on the role of timbre in causing such segregation. The previous study had indicated that *components* within complex tones could stream with components of sequentially presented complexes. This type of component streaming is again

indicated in the present experiments based on the perception of melodic fragments segregated from other components of the stimulus. (Such "melodies on partials" were also reported by McAdams, 1984 b, p. 125).

While components of a harmonic complex generally fuse to form a synthetically-perceived tone, they can be made audible via streaming with a "captor" tone presented sequentially (Dannenbring and Bregman, 1978). The widely-cited analytic experiment of Plomp (1964) appears to have inadvertently provided such a captor in the form of the "probe" tone that was always available to subjects for comparison.

Frequency and temporal proximity are powerful cues in the formation of streams. The repetition of a spectral component within a small time window leads to perceptual streaming of the repeated component. Similar groups can be formed for components close in frequency. It is suggested that the segregation of the mistuned harmonic in our experiments was facilitated by two factors: the lack of harmonicity **within** the complex, coupled with the streaming of matching components **across** the two sounds of the pair. The remaining unmatched component was thus heard as being "split" from the rest of the complex.

An experiment reported by Viemeister and Bacon (1982) also demonstrated the improved audibility of a component within a complex tone given the sequential context of a previous complex **lacking** the target component. They attributed this **"enhancement effect"** to a frequency-specific adaptation process resulting from prior exposure to

the complex with the deleted component. The lack of stimulation of the frequency region corresponding to the deleted component is assumed to heighten its sensitivity relative to that of the adapted regions. When the missing component is included in the second presentation of the complex, it therefore appears to be louder and is heard out from the rest of the complex.

A physiological explanation similar to the one given above for the enhancment effect has also been proposed for stream segregation (van Noorden, 1975; Anstis and Saida, 1985). Van Noorden suggested that "overlapping groups of hair cells have to be excited if temporal coherence is to be heard" (p. 21). ("Temporal coherence" refers to the cohering or streaming together of sounds presented sequentially).

The enhancement observed by Viemeister and Bacon may thus alternatively be interpreted in terms of streaming of the common components between the two complexes. Since the target component did not have a matching component to stream with, it was heard out or "enhanced", as indeed it was in the similar stimuli used in experiment 3.

6.7 Perceptual organization of sound - simultaneous and sequential grouping

When presented a series of sounds juxtaposed in time, a listener is faced with the multiple tasks of organizing information simultaneously as well as sequentially. Simultaneous organization yields the fusion of

components into entities and their perceptual separation from other concurrently-present entities. Sequential organization involves the tracking of groups of components or entities over time.

Bregman and Pinker (1978, p. 19) suggest that "the auditory system must use temporally extended information ... to recover separate descriptions of the several sources" whose combined contributions comprise a complex waveform. They describe two aspects of the acoustic factoring process that enables such parsing: (1) "the grouping together of all the simultaneous frequency components that emanate from a single source at a given moment", and (2) "the connecting over time of the changing frequencies that a single source produces from one moment to the next".

Factors that promote grouping of auditory material act in a competitive way. In some cases, heuristics for both simultaneous and sequential grouping may collaborate or compete with each other in dictating organization (Bregman, 1978; Bregman and Doehring, 1984). An element can be captured out of a sequential grouping by being given a better sound to group with. (Bregman, 1990, p. 651). There is collaboration as well as competition. The experiments reported here add to the growing body of evidence for this dynamic point of view.

6.8 Grouping and feature assignment

A steady-state harmonic complex tone can be perceptually

analyzed into its component tones by careful listening for values of harmonic number up to n ≈5 (Plomp, 1976). Listeners can "hear out" the partials and identify their (spectral) pitches. However, when the partials are coherently modulated in frequency, (i.e. all components are modulated by the same proportion $\Delta f/f$, so as to maintain their harmonicity), listeners are no longer able to hear out the components separately. Rather, the complex is heard as being fused with a unitary virtual pitch (that may, however, be perceived as fluctuating, as in "vibrato"). This observation is interpreted by McAdams as supporting the notion that **grouping processes** (that lead to perceptual fusion) influence the assignment of perceptual features such as pitch.

Earlier work by Bregman and Pinker (1978) also indicated that features such as "timbre" emerge as the **property of a group** of acoustic elements, once those elements have been grouped together. The same idea was stated as a "definition" for timbre by Seashore (1938) more than half a century ago ! He referred to **"timbre"** as that aspect of tone quality "which is **the simultaneous presence or fusion of the fundamental and its overtones at a given moment**" (p. 95).

Darwin (1981) also presented a similar idea for speech sounds. His experiments have shown that spectral components are first compared to yield plausible auditory groups, before phonemic identies are assigned to them.

The many experiments of Bregman and his colleagues (summarized by Bregman, 1990) and the experiments reported here also show

evidence for feature assignment **after** a complex acoustic input has been sorted into coherent groups.

In the present experiments, the influence of context and interactions of components within and across sounds were repeatedly encountered. The spectral loci defined in the first experiment did not acquire a "hard and fast" timbral identity by virtue of the spectral composition. Rather, the timbre was perceived in the context of neighboring components of other sounds. In the inharmonic stimuli of the third experiment, a segregated component was perceived with a pure-tone-like timbre against a backdrop of a low complex tone, when the frequency deviation exceeded the threshold for segregation. Timbres were thus assigned to the two emergent entities *post facto*.

However, evidence also exists to show that perceptual features in turn influence the grouping of auditory events in the larger context of a temporal sequence. S equences comprising tones far apart in frequency are perceptually segregated into streams on the basis of perceived pitch differences. The earlier study on the role of timbre in stream segregation also showed how perceptual groups were formed on the basis of perceived similarity in timbre (Singh, 1987). However, the timbres assigned to the complex tones were susceptible to context effects. Thus, for sequences with tones comprising spectral loci that were near in frequency, interaction between components led to changes in the perceived timbre of the tones, in addition to causing confounding changes in pitch. While timbre served as a cue for perceptual grouping, it also

appeared to be assigned to a sound in the context of the surrounding spectra. This paradox is summed up in the observation by Bregman (1990) that "timbre influences scene analysis, but scene analysis creates timbre" (p. 488).

It is apparent that there exists a rather complex relation between grouping of acoustic elements into an "image" and the derivation of qualities such as the pitch and timbre of that image. The sorting of components into groups and the derivation of these qualities are inherently dependent on the frequency-analysis capability of the auditory system and its ability to discriminate spectral changes.

McAdams has proposed that "sequential organization is based on a context-dependent criterion of spectral continuity" (p. 187). Rather than claiming what comes first; --- the group, or its qualitative features, --- McAdams suggests that both "quality derivation and grouping" operations may be "concurrent" (p. 206). A similar statement made by Hirsh (1974, p. 255) is perhaps the best characterization of the dynamic complexity and marvel of the auditory system:

> "A stream flows in already established channels; and channels are formed by the flowing of streams. It is suggested that both of these observations from nature are evidenced in auditory perception"

Chapter Seven

LEADING NOTES AND RESOLUTIONS

The experiments reported in this dissertation ventured into a region of psychological acoustics that is often left out of mainstream research because of inherent problems in studying perceptual phenomena. Perceptual aspects of complex sounds, unaided by the phonetic labels of speech or tonal ("key") structures of music are difficult to describe. Yet, basic research on audition often *intentionally* avoids the convenience of such referential frameworks, in order to better study "primitive" aspects of hearing with a minimal influence of cognitive "schemas" (Bregman, 1990).

While the perception of relations and criteria for grouping of complex sounds can be made transparent by judicious choice of stimuli and experimental design, the underlying physiological bases and functional mechanisms that accomplish the task of grouping and feature assignment, are harder to elaborate. A dichotomy in experimental approaches has thus emerged in the study of audition. Studies of **perception** focus on describing subjective attributes of sounds and their inter-relations as perceived by listeners. Studies of **sensory psychophysics** on the other hand, focus on sensory limits, mechanisms responsible for coding of stimulus dimensions, and performance of listeners on tasks such as detection, discrimination or identification, without regard to the perceptual cues facilitating the task.

Some efforts to bridge the gap between "perception and psychophysics" have recently emerged. The work on comodulation release of masking (CMR) is a promising start in this area. The obvious, yet usually neglected connection between masking and fusion was mentioned earlier in chapter 2 and chapter 4. The CMR studies started out investigating mechanisms underlying masking, and inadvertently unveiled the related phenomena of fusion and segregation !

A lot of work needs to be done in this "bridge" domain relating perception and traditional psychophysics. Listeners used as subjects in auditory experimentation ultimately rely on **subjective impressions** of perceptual aspects of the stimuli in making judgments. Depending on the degree of training and exposure to different acoustic environments, listeners may use different strategies to relate their percepts to stimuli. The more understanding we acquire about the percepts and their dynamic, context-sensitive nature, the more easily we will be able to relate them to facts about the hearing mechanism obtained from more basic sensory psychophysics. The challenge lies in seeing the connections between the two approaches and being able to bring them into symbiotic alignment in explaining **what** the auditory system listens to, and **how** and **why** it does so.

The present set of experiments endeavored to provide tools and data to establish that this *can* be done. The frequency domain was systematically quantized and the range of associated perceptual changes noted. However, the thrust of the study remained largely descriptive in

nature. This is not necessarily a limitation. Before theories can be constructed to explain percepts, the percepts themselves need to be well explored and understood.

While an attempt has been made to portray the questions asked and the results obtained in the light of both perceptual and psychophysical data on processing of complex sounds, continued experimentation will be necessary to quantify the underlying mechanisms and develop models that can adequately relate percepts associated with spectral changes in complex tones to these mechanisms.

In the present experiments, the stimuli were constrained to be "steady state" at the design stage. The dynamic modulations that emerged as a result of inharmonicity were not investigated in great detail in the temporal domain. Future studies with controlled variation of the waveform would be helpful in the search for principles of fusion in the time domain.

The stimuli in the current experiments were also designed to have flat spectral envelopes, with all components assigned equal amplitudes. This is not a natural spectral design for tonal stimuli. The spectra of many instruments exhibit a "formant structure" with resonances and regions of diminished amplitude. The characteristic pattern of peaks and valleys is the bearer of timbral identity of many instruments, and should thus be investigated more carefully in discrimination experiments of this type.

Deliberate restraint was shown in using varied spectral

envelope shapes in the present experiments in order to avoid potentially confounding relations between frequency changes and amplitudes of the changed components. Timbre changes ensuing from simultaneous variation of these two features would be difficult to attribute to either one. One way to approach this potentially confounding situation, might be to design stimuli of the type used in experiments on "profile analysis" (Green, 1988). One or more components of a complex could be selected for systematic variation in amplitude and timbral consequences of the change, with or without other concurrent changes in the spectrum could be studied.

Similar extensions of this work from a more musical and cognitive viewpoint would also provide a bigger picture of the functions served by perceptual relations between sounds in the real world. Referential systems such as "scales" and "modes" have a profound impact on the discrimination and identification of melodies (Dowling and Harwood,1986). Melodic context has a bearing on judgments of pitch and timbre change within a sequence of sounds (Iverson and Krumhansl, 1989). The rhythmic structure of a sequence also, can influence judgments of pitch changes at strategic locations within the sequence (Jones, Boltz and Kidd, 1982).

Music and speech are two classes of sound structures that have evolved in relation to properties of the auditory system. As our understanding of these structures and their usage grows, so will our understanding of the destination for which they are intended . . .

BIBLIOGRAPHY

ASA (1960) "American standard acoustical terminology (Including mechanical shock and vibration) S1.1", American Standards Assoc. Inc. (since renamed "ANSI S1.1-1960 (R1976): USA standard acoustical terminology"), ANSI, New York.

Bachem, A. (1950) "Tone height and tone chroma as two different pitch qualities", Acta Psychologica, **7**, 80-88.

Balzano, G. J. (1986) "What are musical pitch and timbre ?", Mus. Perc., **3**, 297-314.

Beauchamp, J. W. (1975) "Analysis and synthesis of cornet tones using non-linear interharmonic relationships", J. Aud. Eng. Soc., **23**, 778-795.

Beerends, J. G. (1989) "Pitches of simultaneous complex tones", Ph. D. dissertation, IPO, Eindhoven, Holland.

von Bekesy, G. (1960) "Experiments on hearing", McGraw Hill, New York.

Benade, A. H. (1976) "Fundamentals of Musical Acoustics", Oxford Univ. Press, New York.

Benade, A. H. (1981) "Spectral similarities of tones", J. Acoust. Soc. Am., abstract S37, 69.

Benade, A. H. (1983) "From instrument to ear in a room: Direct or via recording", preprint from the October meeting of the Audio Engineering Society.

Benade, A. H. (1986) "Generic spectrum envelope functions for orchestral wind instruments", J. Acoust. Soc. Am., Suppl. 1 **79**, S93 (reprint of paper and discussion following).

Berger, K. W. (1964) "Some factors in the recognition of timbre", J. Acoust. Soc. Am., **36**, 1888-1891.

Bilsen, F. A. and Ritsma, R. J. (1970) "Repetition pitch and its implication for hearing theory", Acustica, **22**, 63-73.

von Bismarck, G. (1974 a) "Timbre of steady sounds: A factorial investigation of its verrbal attributes", Acustica, **30**, 146-159.

von Bismarck, G. (1974 b) "Sharpness as an attribute of the timbre of steady sounds", Acustica, **30**, 159-172.

Bjørklund, A. (1961) "Analysis of soprano voices", J. Acoust. Soc. Am., 33, 575-582.

de Boer, E. (1956) "On the "residue" in hearing", Academic thesis, Amsterdam, Holland (cited by de Boer (1976))

de Boer, E. (1976) "On the "residue" and auditory pitch perception", Chap. 13 in "Handbook of sensory physiology", vol. V/3, W. D. Keidel and W. D. Neff (eds.), Springer Verlag, Vienna, 479-583.

Bolt, R. (1948) "Wanted - The Formant: Dead or alive", J. Acoust. Soc. Am., **20**(1), 66.

Bregman, A. S. (1978) "The formation of auditory streams", in "Attention and performance" vol. VII, J. Requin (ed.), Lawrence Earlbaum, Hillsdale, New Jersey.

Bregman A. S. (1990) "Auditory scene analysis: The perceptual

organization of sound", MIT press, Cambridge, Mass. .

Bregman A. S. and Campbell, J. (1971) "Primary auditory stream segregation and the perception of order in rapid sequences of tones", J. Exp. Psych., **89**, 244-249.

Bregman A. S. and Pinker, S. (1978) "Auditory streaming and the building of timbre", Can. J. Psych., **32**, 19-31.

Bregman A. S. and Doehring, P. (1984) "Fusion of simultaneous tonal glides: The role of parallelness and simple frequency relations", Perc. and Psychophys., **36**, 251-256.

Bregman A. S., Abramson, J., Doehring, P. and Darwin, C. J. (1985) "Spectral integration based on common amplitude modulation", Perc. and Psychophys., **37**, 483-493.

Bregman, A. S., Levitan, R. and Liao, C. (1990) "Fusion of auditory components: Effects of the frequency of amplitude modulation", Perc. and Psychophys., **47**(1), 68-73.

Broadbent, D. E. and Ladefoged, P. (1957) "On the fusion of sounds reaching different sense organs", J. Acoust. Soc. Am., **29**, 708-710.

Brokx, J. P. L. and Noteboom, S. G. (1982) "Intonation and the perceptual separation of simultaneous voices", J. Phon. **10**, 23-26.

Brugge, J. F., Anderson, D. J., Hind, J. E., and Rose, J. E. (1969) "Time structure of discharges in single auditory-nerve fibers of the squirrel monkey in response to complex periodic sounds", J. Neurophysiol., **32**, 386-401.

Buunen, T. J. F., Festen, J. M., Bilsen, F. A. and van den Brink, G. (1974) "Phase effects in a three-component signal", J. Acoust. Soc. Am., **55**, 297-303.

Buus, S. (1985) "Release from masking caused by envelope fluctuations", J. Acoust. Soc. Am., **78**, 1958-1965.

Cabot, R. C., Mino, M. G., Dorans, D. A., Tackel, I. S. and Breed, H. E. (1976) "Detection of phase shifts in harmonically related tones", J. Aud. Eng. Soc., **24**, 568-571.

Carlson, R., Fant, G. and Granstrom, B. (1975) "Two-formant models, pitch and vowel perception", in "Auditory analysis and speech perception", G. Fant and M. A. A. Tatham (eds.), Academic press, London, 55-82.

Chalikia, M. H. and Bregman, A. S. (1989) "The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation", Perc. and Psychophys., **46**(5), 487-496.

Charbonneau, G. R. (1981) "Timbre and the perceptual effects of three types of data reduction", Comp. Mus. J., **5**(2), 10-19.

Cherry, E. C. (1953) "Some experiments on the recognition of speech with one and with two ears", J. Acoust. Soc. Am., **25**, 975-979.

Clark, M., Robertson, P. and Luce, D. (1964) "A preliminary experiment on the perceptual basis for instrument families", J. Aud. Eng. Soc., **12**, 194-203.

Cohen, E. A. (1980) "The influence of nonharmonic partials on tone perception", Ph. D. dissertation, Stanford University,

Stanford, California.

Cross, D. and Lane, H. (1963) "Attention to single stimulus properties in the identification of complex tones", Univ. Michigan ORA rep. 05613-1-P.

Crowder, R. G. (1989) "Imagery for musical timbre", J. Exp. Psych. (Hum. Perc. and Perf.), **15**(3), 472-478.

Culver, C. (1956) "Musical Acoustics", (4th ed.), McGraw Hill, NY.

Cutting, J. E. (1976) "Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening", Psych. Rev., **83**, 114-140.

Dannenbring, G. L. and Bregman, A. S. (1976) "Stream segregation and the illusion of overlap", J. Exp. Psych. (Hum. Perc. and Perf.), **2**, 544-555.

Dannenbring, G. L. and Bregman, A. S. (1978) "Streaming versus fusion of sinusoidal components of complex tones", Perc. and Psychophys., **24**, 369-376.

Darwin, C. J. (1981) "Perceptual grouping of speech components differing in fundamental frequency and onset-time", Q. J. Exp. Psych., **33A**, 185-207.

Darwin, C. J. (1984) "Perceiving vowels in the presence of another sound: Constraints on formant perception", J. Acoust. Soc. Am., **76**, 1636-1647.

Davis, H., Silverman, S. R. and McAuliffe, D. R. (1951) "Some observations on pitch and frequency", J. Acoust. Soc. Am.,

23, 40-42.

Deutsch, D. (1972) "Octave generalization and tune recognition", Perc. and Psychophys., **11**, 411-412.

Deutsch, D. (ed.) (1982) "The psychology of music", Academic Press, New York.

Dowling, W. J. (1968) "Rhythmic fission and perceptual organization", J. Acoust. Soc. Am., **44**, 369.

Dowling, W. J. and Harwood, D. L. (1986) "Music cognition", Academic Press, New York.

Duifhuis, H., Willems, L. F. and Sluyter, R. J. (1982) "Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception", J. Acoust. Soc. Am., **71** (6), 1568-1580.

Durrant, J. D. and Lovrinic, J. H. (1984) "Bases of hearing science", 2nd ed., Williams and Wilkins, Baltimore/London.

Eagleson, H. W. and Eagleson, O. W. (1947) "Identification of musical instruments when heard directly and over a public address system", J. Acoust. Soc. Am., **19**, 338-342.

Ehresman, D. and Wessel, D. L. (1978) "Perception of timbral analogies"", IRCAM Rep., 13/78, Centre Georges Pompidou, Paris, France.

Elliott, L. L. (1962) "Backward and forward masking of probe tones of different frequencies", J. Acoust. Soc. Am., **34**, 1116-1117.

Erickson, R. (1975) "Sound structure in music", Univ. California Press, Berkeley, California.

Evans, E. F. (1978) "Place and time coding of frequency in the peripheral auditory system: Some physiological pros and cons", Audiol., **17**, 369-420.

Evans, E. F. and Wilson, J. P. (1974) (Eds.) "Psychophysics and physiology of hearing", Academic Press, New York.

Fant, G. (1973) "Speech sounds and features", MIT press, Cambridge, Mass.

Faulkner, A. (1985) "Pitch discrimination of complex signals: Residue pitch or multiple component discriminations ?", J. Acoust. Soc. Am., **78** (6), 1993-2004.

Fastl, H. and Weinberger, M. (1981) "Frequency discrimination for pure and complex tones", Acustica, **49**, 77-78.

Fischler, H. (1967) "Model of the secondary "residue" effect in the perception of complex tones", J. Acoust. Soc. Am., **42**, 759-764.

Flanagan, J. L. (1955) "A difference limen for vowel formant frequency", J. Acoust. Soc. Am., **27**(3), 613-617.

Flanagan, J. L. and Golden, R. M. (1966) "Phase vocoder", Bell Sys. Tech. J., 45, 1493-1509.

Flanagan, J. L. and Guttman, N. (1960a) "On the pitch of periodic pulses", J. Acoust. Soc. Am., **32**, 1308-1319.

Flanagan, J. L. and Guttman, N. (1960b) "Pitch of periodic pulses without fundamental component", J. Acoust. Soc. Am., **32**, 1319-1328.

Flanagan, J. L. and Saslow, M. G. (1958) "Pitch discrimination for

synthetic vowels", J. Acoust. Soc. Am., **30**(5), 435-442.

Fletcher, H. (1924) "The physical criterion for determining the pitch of a musical tone", Phys. Rev., **23**, 427-437.

Fletcher, H. (1934) "Loudness, pitch and timbre of musical tones and their relation to the intensity, the frequency, and the overtone structure", J. Acoust. Soc. Am., **6**(2), 67

Fletcher, H. and Sanders, L. C. (1967) "Quality of violin vibrato tones", J. Acoust. Soc. Am., **41**, 1534-1544.

Fletcher, H., Blackham, E. D. and Geersten, O. N. (1965) "Quality of violin, viola, 'cello and bass-viol tones: I", J. Acoust. Soc. Am., **37**, 851-863.

Forte, A. (1973) "The structure of atonal music", Yale University Press, New Haven, Conn.

Fourier, J. B. J. (1822/1878) "The analytical theory of heat", English transl., A. Freeman (ed.), Cambridge

Gagne, J. P. and Zurek, P. M. (1988) "Resonance frequency discrimination", J. Acoust. Soc. Am., **83**(6), 2293-2299.

Garner, W. R. and Gottwald, R. L. (1968) "The perception and learning of temporal patterns", Quarterly J. Exp. Psychol., **20**, 97-109.

Gerson, A. and Goldstein, J. L. (1978) "Evidence for a general template in central optimal processing for pitch of complex tones", J. Acoust. Soc. Am., 63(2), 498-510.

Goldstein, J. L. (1966) "An investigation of monaural phase perception", Ph. D. Dissertation, University of Rochester, New York (cited by

McAdams, 1984 b).

Goldstein, J. L. (1967) "Auditory spectral filtering and monaural phase perception", J. Acoust. Soc. Am., **41**, 458-479.

Goldstein, J. L. (1973) "An optimum processor theory for the central formation of the pitch of complex tones", J. Acoust. Soc. Am., **54**, 1496-1516.

Goldstein, J. L. (1978) "Mechanisms of signal analysis and pattern perception in periodicity pitch", Audiol., 17, 421-445.

Grandori, F. (1984) "Theoretical and experimental analysis of a central optimal processor for pitch of multicomponent inharmonic tones", Hear. Res., 15, 151-158.

Green, D. M. (1983) "Profile analysis; a different view of auditory intensity discrimination", Am. Psychol., **38**, 133-142.

Green, D. M. (1988 a) "Profile analysis: Auditory intensity discrimination", Oxford psychology series no. 13, Oxford Univ. Press, New York, Oxford.

Green, D. M. (1988 b) "Audition: Psychophysics and perception", Chap. 6 in "Stevens' handbook of experimental psychology" (2nd ed.), vol. 1: Perception and motivation", Atkinson et al. (eds.), John Wiley, New York.

Green, D. M., Kidd, G. and Picardi, M. C. (1983) "Successive versus simultaneous comparisons in auditory intensity discrimination", J. Acoust. Soc. Am., 73, 639-643.

Green, D. M., Mason, C. R. and Kidd, G. (1984) "Profile analysis: Critical

bands and duration", J. Acoust. Soc. Am., **75**, 1163-1167.

Green, D. M. and Swets, J. A. (1974) "Signal detection theory and psychophysics", Krieger, New York.

Grey, J. M. (1975) "An exploration of musical timbre", Ph. D. Dissertation, Stanford University, Stanford, California (Dept. of Music Tech. Rep. No. STAN-M-2)

Grey, J. M. (1977) "Multidimensional perceptual scaling of musical timbres", J. Acoust. Soc. Am., **61**, 1270-1277.

Grey, J. M. (1978) "Timbre discrimination in musical patterns", J. Acoust. Soc. Am., **64**, 467-472.

Grey, J. M. and Gordon, J. W. (1978) "Perceptual effect of spectral modifications in musical timbres", J. Acoust. Soc. Am., **63**, 1493-1500.

Grey, J. M. and Moorer, J. A. (1977) "Perceptual evaluation of synthesized musical instrument tones", J. Acoust. Soc. Am., **62**, 454-462.

Hall, J. W. (1987) "Experiments on comodulation masking release", Chap. 6 in Yost, W. A. and Watson, C. S. (Eds.) "Auditory processing of complex sounds", Lawrence Earlbaum, Hillsdale, New Jersey, 57-66.

Hall, J. W. and Grose, J. H. (1990) "Comodulation masking release and auditory grouping", J. Acoust. Soc. Am., **88** (1), 119-125.

Hall, J. W. and Soderquist, D. R. (1975) "Encoding and pitch strength of complex tones", J. Acoust. Soc. Am., **58**, 1257-1261.

Hall, J. W., Grose, J. H. and Haggard, M. P. (1988) "Comodulation masking release for multicomponent signals", J. Acoust. Soc. Am., **83**(2), 677-686.

Hall, J. W., Haggard, M. P. and Fernandes, M. A. (1984) "Detection in noise by spectro-temporal pattern analysis", J. Acoust. Soc. Am., **76**, 50-56.

Handel, S. (1974) "Perceiving melodic and rhythmic auditory patterns", J. Exp. Psychol., **103**, 922-933.

Handel, S. (1989) "Listening: An introduction to the perception of auditory events", MIT press, Cambridge, Mass.

Harre, R. and Lamb, R. (eds.) (1983) "The encylopedic dictionary of psychology", (definition of "perception"; p. 450-451), MIT Press, Cambridge, Mass.

Harris, J. D. (1952) "Pitch discrimination", J. Acoust. Soc. Am., **24**, 750-755.

Hartmann, W. M. (1988) "Pitch perception and the segregation and integration of auditory entities", Chap. 21 in "Auditory function", G. M. Edelman, W. E. Gall and W. M. Cowan (eds.), John Wiley and Sons, Inc., New York.

Hartmann, W. M. (1989) Personal communication.

Hartmann, W. M., McAdams, S., Gerzso, A. and Boulez, P. (1986) "Discrimination of spectral density", J. Acoust. Soc. Am., **79**(6), 1915-1925.

Hawks, J. W. (1990) "Perceptual aspects of a vowel space", Ph. D.

dissertation, Washington University, St. Louis, Missouri.

Helmholtz, H. L. F., von (1877/1954) "On the sensations of tone as a physiological basis for the theory of music", English translation by A. J. Ellis, Dover, New York.

Henning, G. B. and Grosberg, S. L. (1968) "Effect of harmonic components on frequency discrimination", J. Acoust. Soc. Am., **44**, 1386-1389.

Hesse, H. P. (1982) "The judgment of musical intervals", Chapter 11 in "Music, mind and brain: The neuropsychology of music", M. Clynes (ed.), Plenum, New York.

Hind, J. E., Anderson, D. J., Brugge, J. F. and Rose, J. E. (1967) "Coding of information pertaining to paired low-frequency tones in single auditory-nerve fibers of the squirrel monkey", J. Neurophysiol., **30**, 794-816.

Hirsh, I. J. (1959) "Auditory perception of temporal order", J. Acoust. Soc. Am., **31**(6), 759-767.

Hirsh, I. J. (1974) "Temporal order and auditory perception", in "Sensation and measurement", H. R. Moskowitz et al. (eds.), D. Reidel, 251-258.

Hirsh, I. J. (1988) "Auditory perception and speech", Chap. 7 in "Stevens' handbook of experimental psychology" (2nd ed.), vol. 1: "Perception and motivation", Atkinson et al. (eds.), John Wiley, New York.

Houtsma, A. J. M. (1979) "Musical pitch of two-tone complexes and

predictions by modern pitch theories", J. Acoust. Soc. Am., **66**, 87-99.

Houtsma, A. J. M. and Canning, J. M. (1983) "Pitch perception of simultaneous complex tones", Ann. Prog. Rep., IPO, Eindhoven, Holland.

Houtsma, A. J. M. and Goldstein, J. L. (1972) "The central origin of the pitch of complex tones: Evidence from musical interval recognition", J. Acoust. Soc. Am., **51**, 520-529.

Houtsma, A. J. M. and Smurzynski, J. (1990) "Pitch identification and discrimination for complex tones with many harmonics", J. Acoust. Soc. Am., 87(1), 304-310.

Huggins, W. H. (1952) "A phase principle for complex frequency analysis and its implications in auditory theory", J. Acoust. Soc. Am., **24**, 582-589.

Iverson, P. and Krumhansl, C. (1989) "Pitch and timbre interaction in isolated tones and in sequences", J. Acoust. Soc. Am., Suppl. 1, **86**, S58.

Jesteadt, W. and Norton, S. (1985) "The role of suppression in psychophysical measures of frequency selectivity", J. Acoust. Soc. Am., **78**(1), 365-374.

Jones, M. R. (1976) "Time, our lost dimension: Toward a new theory of perception, attention, and memory", Psych. Rev., **82**, 323-355.

Jones, M. R., Boltz, M. and Kidd, G. (1982) "Controlled attending as a

function of melodic and temporal context", Perc. and Psychophys., **32**(3), 211-218.

Kendall, R. A. (1986) "The role of acoustical signal partitions in listener categorization of musical phrases", Music Perc., **4** (2), 85-214.

Koffka, K. (1935) "Principles of gestalt psychology", Harcourt and Brace, New York.

Kohlrausch, A. and Jacobi, G. (1989) "Learning effects in a simple masking experiment using complex-tone maskers", Assoc. Res. Otolaryngol., Conf. Proc., 258-259.

Krumhansl, C. (1979) "The psychological representation of musical pitch in a tonal context", Cog. Psych., 11, 346-374.

Krumhansl, C. and Shepard, R. N. (1979) "Quantification of the hierarchy of tonal functions within a diatonic context", J. Exp. Psych: Hum. Perc. and Perf., 5, 579-594.

Lai, W. K., Engebretson, A. M., Metzger, M. and Singh, P. G. (1987) "PSYACX - a program for psychoacoustic experimentation", software documentation, Central Institute for the Deaf, St. Louis, Missouri.

Lauter, J. L. (1983) "Stimulus characteristics and relative ear advantages: A new look at old data", J. Acoust. Soc. Am., **74**(1), 1-17.

Lauter, J. L. (1985) "Individual differences in the perception of frequency changes in three-element sequences", J. Acoust. Soc.

Am., (Suppl.1), **77**, S36.

Lauter, J. L. and Singh, P. (1985) "Melody perception", Per. Prog. Rep. #28, Research Dept., Central Institute for the Deaf, St. Louis, Missouri.

Lichte, W. H. (1941) "Attributes of complex tones", J. Exp. Psychol., **28**, 455-480.

Licklider, J. C. R. (1954) "'Periodicity' pitch and 'place' pitch", J. Acoust. Soc. Am., **26**, 945 (A).

Luce, D. A. (1975) "Dynamic spectrum changes of orchestral instruments", J. Aud. Eng. Soc., **23**(7), 565-568.

Makhoul, J. (1975) "Linear prediction: A tutorial review", Proc. IEEE, **63**, 561-580.

Mathes, R. C. and Miller, R. L. (1947) "Phase effects in monaural perception", J. Acoust. Soc. Am., **19**, 780-797.

Mathews, M. V. and Pierce, J. R. (1980) "Harmony and nonharmonic partials", J. Acoust. Soc. Am., **68**, 1252-1257.

McAdams, S. (1984 a) "The auditory image: A metaphor for musical and psychological research on auditory organization", in "Cognitive processes in the perceptios of art", R. Crozier and A. Chapman (eds.), North-Holland publ. , Amsterdam, Holland.

McAdams, S. (1984 b) "Spectral fusion, spectral parsing and the formation of auditory images", Ph. D. Dissertation, Stanford University, Stanford, California (Dept. of Music Tech. Rep. No. STAN-M-22)

McAdams, S. and Bregman, A. S. (1979) "Hearing musical streams", Comp. Mus. J., **3**(4), 26-43.

McClelland, K. D. and Brandt, J. F. (1969) "Pitch of frequency-modulated sinusoids", J. Acoust. Soc. Am., **45**, 1489-1498.

McFadden, D. (1986) "Comodulation masking release: Effects of varying the level, duration, and time delay of the cue band", J. Acoust. Soc. Am., **80**, 1658-1667.

Mermelstein, P. (1978) "Difference limens for formant frequencies of steady-state and consonant-bound vowels", J. Acoust. Soc. Am., **63**(2), 572-580.

Miller, D. C. (1926) "Science of musical sounds", MacMillan, New York.

Miller, G. A. and Taylor, W. G. (1948) "The perception of repeated bursts of noise", J. Acoust. Soc. Am., **20**, 171-180.

Miller, J. D. (1984) "Auditory processing of the acoustic patterns of speech", Arch. Otolaryng., 110, 154-159.

Miller, J. D. (1989) "Auditory perceptual interpretation of the vowel", J. Acoust. Soc. Am., **85**, 2114-2134.

Miller, J. R. and Carterette, E. C. (1975) "Perceptual space for musical structures", J. Acoust. Soc. Am., **58**, 711-720.

Monahan, C. B. and Carterette, E. (1985) "Pitch and duration as determinants of musical space", Mus. Perc., **3**, 1-32.

Moore, B. C. J. (1982) "An introduction to the psychology of hearing", (2nd ed.), Academic Press, London.

Moore, B. C. J. (1987) "The perception of inharmonic complex tones",

Chap. 17 in Yost, W. A. and Watson, C. S. (Eds.) "Auditory processing of complex sounds", Lawrence Earlbaum, Hillsdale, New Jersey.

Moore, B. C. J. and Glasberg, B. R. (1990) "Frequency discrimination of complex tones with overlapping and non-overlapping harmonics", J. Acoust. Soc. Am., **87**(5), 2163-2177.

Moore, B. C. J., Glasberg, B. R. and Shailer, M. J. (1984) "Frequency and intensity difference limens for harmonics within complex tones", J. Acoust. Soc. Am., **75**(2), 550-561.

Moore, B. C. J., Glasberg, B. R. and Peters, R. W. (1985 a) "Relative dominance of individual partials in determining the pitch of complex tones", J. Acoust. Soc. Am., **77**(5), 1853-1860.

Moore, B. C. J., Peters, R. W. and Glasberg, B. R. (1985 b) "Thresholds for the detection of inharmonicity in complex tones", J. Acoust. Soc. Am., **77**(5), 1861-1867.

Moore, B. C. J., Glasberg, B. R. and Peters, R. W. (1986) "Thresholds for hearing mistuned partials as separate tones in harmonic complexes", J. Acoust. Soc. Am., **80**(2), 479-483.

Moorer, J. A. (1977) "Signal processing aspects of computer music: a survey", Proc. IEEE, **65**, 1108-1137.

van Noorden, L. P. A. S. (1971) "Rhythmic fission as a function of tone rate", IPO Ann. Prog. Rep., Eindhoven, Holland.

van Noorden, L. P. A. S. (1975) "Temporal coherence in the perception of tone sequences", Ph. D. dissertation, Tech. Hogeschool,

Eindhoven, Holland.

Nordmark, J. O. (1978) "Frequency and periodicity analysis", Chap. 7 in "Handbook of perception, vol. IV - Hearing", Carterette, E. and Friedman, M. P. (eds.), Academic Press, New York.

Ohm, G. S. (1843) "Uber die definition des tones, nebst daran geknupfter Theorie der Sirene und ahnlicher tonbildender Vorrichtungen", Ann. Phys. Chem., **59**, 513-565 (cited by de Boer, 1976).

Olson, H. F. (1967) "Music, physics and engineering", Dover, New York.

Parsons, T. W. (1976) "Separation of speech from interfering speech by means of harmonic selection", J. Acoust. Soc. Am., **60**(4), 911-918.

Patterson, R. D. (1973) "The effects of relative phase and number of components on residue pitch", J. Acoust. Soc. Am., **53**, 1565-1572.

Patterson, R. D. (1989) "The tone height of multi-harmonic tones", Proc. of the 1st international conference on "Music perception and cognition", Kyoto, Japan, 119-124.

Patterson, R. D. (1989) "Timbre and tone height", J. Acoust. Soc. Am., Suppl. 1, **86**, S58.

Patterson, R. D. and Wightman, F. L. (1976) "Residue pitch as a function of component spacing", J. Acoust. Soc. Am., **59**, 1450-1459.

Plomp, R. (1964) "The ear as a frequency analyzer", J. Acoust. Soc. Am., **36**, 1628-1636.

Plomp, R. (1967) "Pitch of complex tones", J. Acoust. Soc. Am., 41, 1526-1533.

Plomp, R. (1970) "Timbre as a multidimensional attribute of complex tones", in "Frequency analysis and periodicity detection in hearing", Plomp, R. and Smoorenburg, G. F. (eds.), Sijthoff, Leiden, The Netherlands, 398-414.

Plomp, R. (1976) "Aspects of tone sensation", Academic Press, New York.

Plomp, R. and Levelt, W. J. M. (1965) "Tonal consonance and critical bandwidth", J. Acoust. Soc. Am., 38, 548-560.

Plomp, R. and Mimpen, A. M. (1967) "The ear as a frequency analyzer. II", J. Acoust. Soc. Am., 43(4), 764-767.

Plomp, R. and Smoorenburg, G. F. (eds.) (1970) "Frequency analysis and periodicity detection in hearing", Sijthoff, Leiden, The Netherlands.

Plomp, R. and Steenecken, H. J. M. (1969) "Effect of phase on the timbre of complex tones", J. Acoust. Soc. Am., **46**, 409-421.

Plomp, R. and Steenecken, H. J. M. (1971) "Pitch versus timbre", Proc. 7th Int. Cong. Acous., Budapest, **3**, 377-380.

Pollack, I. (1978) "Decoupling of auditory pitch and stimulus frequency: The Shepard phenomenon revisited", J. Acoust. Soc. Am., **63**(1), 202-206.

Rand, T. C. (1974) "Dichotic release from masking for speech", J. Acoust. Soc. Am., **55**, 678-680.

Rasch, R. (1978) "The perception of simultaneous notes such as in polyphonic music", Acustica, **40**, 21-33.

Revesz, G. (1954) "Introduction to the psychology of music", University of Oklahoma Press, Norman, Okl.

Risset, J. C. (1971) "Paradoxes de hauteur: Le concept de hauteur sonore n'est pas le meme pour tout le monde", Proc. of the 7th ICA, Budapest, 20, S 10, 613-616.

Risset, J. C. (1978) "Hauteur et timbre des sons", dans le Bull. d'audiophonologie, **3**, 9-26, Besancon, France.

Risset, J. C. (1978) "Musical acoustics", Chap. 12 in "Handbook of perception (vol. IV - Hearing)", E. Carterette and M. Friedman (eds.), Academic press, New York.

Risset, J. C. (1986) "Pitch and rhythm paradoxes: Comments on "Auditory paradox based on fractal waveform", [J. Acoust. Soc. Am. 79, 186-189 (1986)]", J. Acoust. Soc. Am., **80**(3), 961-962.

Risset, J. C. and Mathews, M. V. (1969) "Analysis of musical instrument tones", Physics Today, **22**(2), 23-30.

Risset, J. C. and Wessel, D. L. (1982) "Exploration of timbre by analysis and synthesis", Ch. 2. in D. Deutsch (ed.), "The psychology of music", Academic Press, New York.

Ritsma, R. J. (1962) "Existence region of the tonal residue I", J. Acoust. Soc. Am., **34**, 1224-1229.

Ritsma, R. J. (1963) "Existence region of the tonal residue II", J. Acoust. Soc. Am., **34**, 1241-1245.

Ritsma, R. J. (1963) "On pitch discrimination of residue tones", International Audiology, **2**, 34-37.

Ritsma, R. J. (1967) "Frequencies dominant in the perception of the pitch of complex tones", J. Acoust. Soc. Am., **42**, 191-198.

Ritsma, R. J. and Hoekstra, A. (1974) "Frequency selectivity and the tonal residue", in Zwicker, E. and Terhardt, E. (eds.); "Facts and models in hearing", Springer-Verlag, New York, Berlin. , 156-163.

Rose, J. E., Brugge, J. F., Anderson, D. J. and Hind, J. E. (1967) "Phase-locked response to low-frequency tones in single auditory nerve fibres of the squirrel monkey, J. Neurophysiol., **30**, 769-793.

Rutherford, W. (1886) "A new theory of hearing", J. Anat. Physiol., 21, 166-168.

Saldanha, E. L. and Corso, J. F. (1964) "Timbre cues for the recognition of musical instruments", J. Acoust. Soc. Am., **36**, 2021-2026.

Saunders, F. (1946) "Analysis of the tones of a few wind instruments", J. Acoust. Soc. Am., **18**(2), 395-401.

Schaeffer, P. (1966) "Traite des objets musicaux", Paris: Ed. du Seuil.

Scharf, B. (1970) "Critical bands", in "Foundations of modern auditory theory", vol. 1, J. V. Tobias (ed.), Academic Press, New York, 159-202.

Scheffers, M. T. M. (1983) "Sifting vowels: Auditory pitch analysis and sound segregation", Doctoral thesis, University of Groningen, The Netherlands.

Scheffers, M. T. M. (1984) "Discrimination of fundamental frequency of

synthesized vowel sounds in a noise background", J. Acoust. Soc. Am., **76**(2), 428-434.

Schindler, K. W. (1984) "Dynamic timbre control for real-time digital synthesis", Comp. Mus. J., **8**(1), 28-42.

Schmid, C. E. (1977) "Acoustic pattern recognition of musical instruments", Ph. D. dissertation, University of Washington, Seattle, Wash.

Schouten, J. F. (1938) "The perception of subjective tones", Proc. Kon. Ned. Akad. Wetensch., **41**, 1086-1093.

Schouten, J. F. (1940 a) "The residue, a new component in subjective sound analysis", Proc. Kon. Ned. Akad. Wetensch., 43, 356-365.

Schouten, J. F. (1940 b) "The residue and the mechanism of hearing", Proc. Kon. Ned. Akad. Wetensch., **43**, 991-999.

Schouten, J. F., Ritsma, R. J. and Cardozo, B. L. (1962) "Pitch of the residue", J. Acoust. Soc. Am., **34**, 1418-1424.

Schroeder, M. R. (1959) "New results concerning monaural phase sensitivity", J. Acoust. Soc. Am., **31**, 1579.

Schubert, E. D. and Nixon, J. C. (1970) "On the relation between temporal envelope patterns at two different points in the cochlea", Technical Report, Hearing Science Laboratories, Stanford University (cited by McAdams, 1984 b).

Seashore, C. E. (1938) "Psychology of music", McGraw Hill, New York, (reprinted in 1977 by Dover, New York).

Seebeck, A. (1843) "Uber die Sirene", Ann. Phys. Chem., 60, 449-481

(cited by de Boer, 1976). .

Shannon, C. E. and Weaver, W. (1949) "The mathematical theory of communication", Univ. of Illinois press, Urbana.

Shepard, R. N. (1982) "Structural representations of musical pitch", in "The psychology of music", D. Deutsch (ed.), Academic Press, New York.

Shepard, R. N., Romney, A. K. and Nerlove, S. B. (eds.) (1972) "Multidimensional scaling", Seminar press, New York.

Shower, E. G. and Biddulph, R. (1931) "Differential pitch sensitivity of the ear", J. Acoust. Soc. Am., **3**, 275-287.

Siebert, W. M. (1970) "Frequency discrimination in the auditory system: Place or periodicity mechanisms?", Proc. IEEE, 58, 723-730.

Singh, P. G. (1984) "Dimensional tradeoffs in the perception of complex-tone sequences", A.M. Thesis, Washington University, St. Louis, Missouri.

Singh, P. (1985) ""Spectral locus and spectral spacing as determinants in the perceptual organization of complex-tone sequences", J. Acoust. Soc. Am., Suppl. 1 **80**, S40.

Singh, P. G. (1987) "Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre ?", J. Acoust. Soc. Am., **82**(3), 886-899.

Singh, P. G. (1988) "Discrimination of missing F0 and the influence of competing pitch and timbre cues", J. Acoust. Soc. Am., Suppl. 1

84, S143.

Singh, P. G. (1989) "Interaction of timbre and pitch in spectral discrimination tasks using complex tones", J. Acoust. Soc. Am., Suppl. 1 **86**, S58

Slaymaker, F. H. (1970) "Chords from tones having stretched partials", J. Acoust. Soc. Am., **47**, 1569-1571.

Slawson. A. W. (1968) "Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency", J. Acoust. Soc. Am., **43**, 87-101.

Slawson. A. W. (1985) "Sound color", Univ. Calif. press, Berkeley, CA.

Smoorenburg, G. F. (1970) "Pitch perception of two-frequency stimuli", J. Acoust. Soc. Am., **48**, 924-942.

Soderquist, D. R. (1970) "Frequency analysis and the critical band', Psychon. Sc., **21**, 117-119.

Stevens, K. N. (1952) "The perception of vowel formants", J. Acoust. Soc. Am., **24**, 450 (A).

Stevens, S. S. and Davis, H. (1938) "Hearing, its psychology and physiology", John Wiley, New York.

Strong, W. and Clark, M. (1967) "Perturbations of synthetic orchestral wind instrument tones", J. Acoust. Soc. Am., **41**, 277-285.

Terhardt, E. (1974) "On the perception of periodic sound fluctuations (Roughness)", Acustica, **30**, 201-213.

Terhardt, E. (1974) "Pitch, consonance and harmony", J. Acoust. Soc. Am., **55**, 1061-1069.

Terhardt, E. (1979) "Calculating virtual pitch", Hearing Res., **1**, 155-182.

Terhardt, E., Stoll, G. and Seewann, M. (1982 a) "An algorithm for extraction of pitch and pitch salience from complex tonal signals", J. Acoust. Soc. Am., **71**, 679-678.

Terhardt, E., Stoll, G. and Seewann, M. (1982 b) "Pitch of complex tone signals according to virtual pitch theory: Tests, examples and predictions", J. Acoust. Soc. Am., **71**, 671-678.

Thurlow, W. R. and Small, A. M. (1955) "Pitch perception for certain periodic auditory stimuli", J. Acoust. Soc. Am., **27**, 132-137.

Tobias, J. V. (ed.) (1972) "Foundations of modern auditory theory", vol.I, Academic Press, New York.

Toch, E. (1948/1977) "The shaping forces in music: An inquiry into the nature of harmony, melody, counterpoint, form", Dover, New York.

Tolstov, G. P. (1962) "Fourier series", English translation by R. A. Silverman, Dover, New York.

Viemeister, N. F. and Bacon, S. P. (1982) "Forward masking by enhanced components in harmonic complexes", J. Acoust. Soc. Am., **71**(6), 1502-1507.

Viemeister, N. F. and Fantini, D. A. (1987) "Discrimination of frequency ratios", in Yost, W. A. and Watson, C. S. (Eds.) "Auditory processing of complex sounds", Lawrence Earlbaum, Hillsdale, New Jersey, 47-56.

Ward, W. D. (1954) "Subjective musical pitch", J. Acoust. Soc. Am., **26**(3), 369-380.

Warren, R. M, Obusek, C. J., Farmer, R. M. and Warren, R. P. (1969) "Auditory sequence: Confusion of patterns other than speech or music", Science, **164**, 586-587.

Watson, C. S., Kelly, W. J. and Wroton, H. W. (1976) "Factors in the discrimination of tonal patterns. II. Selective attention and memory under various levels of uncertainty", J. Acoust. Soc. Am., **60**, 1176-1186.

Watson, C. S., Wroton, H. W. , Kelly, W. J. and Benbasset, C. A. (1975) "Factors in the discrimination of tonal patterns. I. Component frequency, temporal position and silent intervals", J. Acoust. Soc. Am., **75**, 1175-1185.

Wedin, L. and Goude, G. (1972) "Dimensional analysis of the perception of instrumental timbre", Scand. J. Psych., **13**, 228-240.

Weir, C. C., Jesteadt, W. and Green, D. M. (1977) "Frequency discrimination as a function of frequency and sensation level", J. Acoust. Soc. Am., **61**, 178-184.

Wente, E. C. (1935) "Characteristics of sound transmission in rooms", J. Acoust. Soc. Am., **7**, 123.

Wessel, D. L. (1979) "Timbre space as a musical control structure", Computer Mus. J., **3**(2), 45-52.

Wessel, D. L., Bristow, D. and Settel, Z. (1987) "Control of phrasing and articulation in synthesis", International Computer Music

Conference Proceedings, 108-116.

Wever, E. G. (1949) "Theory of hearing", Dover reprint, New York, 1970.

Wever, E. G. and Bray, C. W. (1930) "Action currents in the auditory nerve in response to acoustic stimulation", Proc. Nat. Acad. Sc., **16**, 344-350.

Whitfield, I. C. (1970) "Central nervous processing in relation to spatio-temporal discrimination of auditory patterns", in "Frequency analysis and periodicity detection in hearing", Plomp, R. and Smoorenburg, G. F. (eds.), Sijthoff, Leiden, The Netherlands, 136-152.

Wightman, F. L. (1973) "The pattern transformation model of pitch", J. Acoust. Soc. Am., **54**, 407-416.

Winckel, F. (1967) "Music, sound and sensation: A modern exposition", Dover, New York.

Yost, W. A. and Watson, C. S. (eds.) (1987) "Auditory processing of complex sounds", Lawrence Earlbaum Assoc., Hillsdale, New Jersey.

Young, R. W. (1960) "Eleven articles related to musical acoustics" reprint from McGraw Hill encyclopedia of Science and Technology, McGraw Hill, New York (cited by Schmid, 1977).

Zwicker, E. (1952) "Die grenzen der Horbarkeit der amplitudenmodulation und der frequenzmodulation eine tones", Acustica, **2** (Beih. 3), 125-133 (cited by Goldstein, 1967).

Zwicker, E. and Scharf, B. (1965) "A model of loudness summation", Psych. Rev., **72**, 3-26.

Zwicker, E. and Terhardt, E. (eds.) (1974) "Facts and models in hearing", Springer-Verlag, New York, Berlin.