

**WATCHING MUSIC: THE EFFECTS OF CONTRAST, CONTEXT AND
FAMILIARITY ON THE VISUAL PERCEPTION OF MUSIC**

by

Patrick Gurr

**THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS**

in the Faculty

of

Education

© Patrick Gurr 1996

SIMON FRASER UNIVERSITY

April 1996

**All rights reserved. This work may not be
reproduced in whole or in part, by photocopy
or other means, without permission of the author.**



National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services Branch

Direction des acquisitions et
des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file *Votre référence*

Our file *Notre référence*

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-16900-6

Canada

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

**Watching Music: The Effects of Contrast, Context and Familiarity on the
Visual Perception of Music**

Author:

(Signature)

Patrick Dana Gurr

(Name)

April 2, 1996

(Date)

APPROVAL

NAME Patrick Dana Gurr

DEGREE Master of Arts

TITLE Watching Music: The Effects of Contrast, Context and Familiarity on the Visual Perception of Music

EXAMINING COMMITTEE:

Chair Michael Manley-Casimir

Robert Walker, Professor
Senior Supervisor

Barry D. Truax, Professor
School of Communication
Member

Dr. Peter Gouzouasis
Dept. of Curriculum Studies
Faculty of Education
University of British Columbia
External Examiner

Date: April 2, 1996

Abstract

The visual pacing and contrast used in audiovideo presentations of a music performance will either reinforce or detract from the performer as they deliver their musical message. The present study has researched visual editing, pacing and contrast on our perception of a musical performance by comparing how four different degrees of audiovisual synchronisation and contrast play a part in the research subjects' response.

In this study, a music performance recorded on both audio and videotape was edited into four different versions where pitch errors were supported with varying degrees of synchronised video. Subjects watched the edited videotapes and their response time to the pitch errors were analysed.

It was found that music perception can be influenced by the recognition of contrasts or discrepancies occurring between the audio and visual aspects of a performance, and on those events separated, but not disassociated from adjacent events. Aural familiarity with the mode and style of presentation of the music also plays a large part in their visual perception of these events.

Recommendations are made for further research, especially in the area of teacher-centred multimedia development and implementation.

Dedication

To Jennifer Anne Holdham, whose wisdom and resolve are my two greatest educational resources.

Acknowledgments

The author is grateful to Marty Hasselbach and students at Columbia Academy for their cooperation in this study. I would also like to express my sincere appreciation to Robert Walker, and Barry Truax for their initial interest, and continued support and intelligent insight into what I hoped to accomplish.

Table of Contents

Approval	ii
Abstract	iii
Dedication	iv
Acknowledgment	v
Table of Contents	vii
1. Introduction	1
- Reasons leading to the Study	7
- The Study	11
- Problems	11
- Limitations	11
2. Related Research	14
- Related Research into the Perception of Music	14
- Related Research on the effects of Visual Perception of Music	31
3. The Experiment	43
- Introduction	43
- Subjects	44
- Auditory Stimuli	45
- Procedure of the Experiment	47
- Audiovisual Stimuli	49

4. Results and Discussion	58
- Experiment #1	59
- Experiment #2	62
- Experiment #3	71
- Comparing the reactions to Experiment #2 and #3	75
- Experiment #4: introduction	76
- Experiment #4a	77
- Comparing the reactions to Experiment #2 and #4a	78
- Experiment #4b	82
- Comparing the reactions to Experiment #3 and #4b	84
5. Conclusions and Recommendations	89
Bibliography	96
Appendices	102

Figures

Figure 1	Students' seating on the day of research	48
Figure 2	Group average reaction times to Experiment #2	65
Figure 3	Group average reaction times to Experiment #3	73
Figure 4	Group average reaction times to Experiment #4a	78
Figure 5	Group average reaction times to Experiment #2 vs. #4a	80
Figure 6.1	Group average reaction times to Experiment #4b	84
Figure 6.2	Group average reaction times to Experiment #2 vs. #4b	85

Tables

Table 1	Elapsed times between edits in Experiment #1	60
Table 2.1	Group average reaction times to Experiment #2	65
Table 2.2	Elapsed times between edits in Experiment #2	66
Table 3.1	Group average reaction times to Experiment #3	73
Table 3.2	Elapsed times between edits in Experiment #3	74
Table 4.1	Group average reaction times to Experiment #4a	78
Table 4.2	Elapsed times between edits in Experiment #4a	78
Table 5	Group average reaction times to Experiment #2 vs. #4a	80
Table 6	F-Test: Two sample for Variances, Experiment #2 vs. #4a	81
Table 7.1	Elapsed times between edits in Experiment #4b	83
Table 7.2	Group average reaction times to Experiment #3 vs. #4b	84

CHAPTER 1

INTRODUCTION

As North America moves toward the '500 channel universe' there seems to be a growing consensus that our computers or televisions will soon interlink us with heretofore inconceivable amounts of information. The mode of communication expected to evolve from this "information highway" could potentially rewrite society's future in much the same manner as Gutenberg's presses did five hundred years ago. The first significant changes have already occurred through the rapid development of interactive computer software, CD ROMs, JPEG , MPEG and other audio-visual information sources. The next change will occur when telecommunication and television broadcasting becomes interactive - where the information source and its recipient can work through a number of interfaces until they find one which suits the preferences of both parties while still allowing for the efficient transmission of the requested information. Still another shift will occur when the information marketplace taps into the potential of "real-time" computer-based conferencing; when both the marketer and client, or teachers and student become linked through camera / computer / television systems, making it possible for them to communicate through text, moving images, sound, and interpersonal conversation simultaneously.

Because, at present, there appears to be no indication of exclusionary practices on the part of broadcasters and telecommunication companies, it is possible to hope that arts education in general - and music education in particular - will benefit in some manner from this expanding technological force. It seems reasonable to assume that at least a portion of the many financial and personnel resources already being spent researching and developing methodologies to teach music in schoolrooms could be diverted to understanding and utilising these emerging technologies.

At the moment, it appears that most media experts and developers are more concerned with the mode of delivery than they are with the content it will deliver. They seem intent on developing faster and more accurate methods to simultaneously transmit and/or receive high quality digital sound and video through the computer, using an Internet-like interface; a mode of communication that may well serve as the paradigm for all future communications growth. The Internet - although not a new technology itself - has led to an unparalleled growth in information processing technologies, including accelerated modems speeds, audio and video digitisers, data compression algorithms, physical linkups via fiber optics, and Internet servers. These Internet servers now include established telephone companies such as B.C.Telephone and cable broadcasters such as Rogers Communications.

Aside from the obvious legal restraints, there are few concerns or restrictions on the part of the servers about what they provide the public through their interface. The content is the client's business, and the marketplace will decide the viability of the information. However, the concentration on the shape and style of the delivery system calls into question the semiotic nature of the information:

Clearly, human-technology relationships are greatly effected not only by technological form, but also by the deliberate allocation of specific tasks to specific technologies. Any technology has specific components, processes, and characteristics that comprise its instrumentation. However, unlike its instrumentation, which is highly visible, a technology's effects are often subtle and, moreover, usually occur as a chain reaction. (Ullman, 1987 p.96.)

The subtle effect of the information technology on the viewer has become of great concern to educators (Lepper,1985; Beentjes,1987; Kerr, 1989; Webb, 1991) knowing as they do, that a product which captures the imagination of their students usually brings about a chain reaction; a string of clones and duplicates of the original idea. The idea of the moment is the World Wide Web (WWW) on the Internet; an interactive information model similar to the information style of CD-ROM technology, which is emerging as a lingua-franca mode of delivery. Normally, a delivery system of this magnitude and potential would undergo months - perhaps years - of study before it is unleashed on our students in such an all-encompassing manner.

However, educators are being dragged along into the Internet under the commercial marketplace's impetus at a time when they are barely able to measure the influence of a single CD-ROM, much less properly analyse how this mode of instruction works in general, or in context. Determining its working in a given context would mean developing a means of testing the reliability and practicality of the Internet as an instructional strategy using a multitude of different modes of delivery. At this point, the fait accompli mode of delivery is the Internet - or something very similar to it - and the means of delivery is multimedia.

Although it is arguable whether educators are adequately prepared to do so, the educational community must begin giving constructive opinions about what this interface - a kind of "music-education television" - should look, interact and "feel" like. However, the main stumbling block to their positive influence on the emerging interface resides in the fact that many music teachers continue to be influenced by their acquired tradition of personal interaction between student and teacher, which is based on a one-to-one relationship, traditionally justified by the belief that certain aspects of music education cannot be easily or simply expressed from afar, requiring, as they do, a tacit confirmation about their value because they belong to the higher order skills, such as aesthetics, creativity, etc. It has been long argued that without these components we are only passing on craft.

Recent advances in technology - especially the computer-driven multimedia technology - may be consciously or unconsciously breaking down the close personal relationship between the teacher and pupil, making it easier for students to receive educational material in their homes. The one-to-one relationship may be further destabilised when it no longer becomes cost-effective for governments to support students to personally attend a post secondary institution of their choice, or when scheduling of private music classes become more difficult due to dwindling finances. Regardless of the reason, it is perceivable that students may soon begin spending a considerable amount of time away from the traditional classroom, achieving their understanding of music through some form of information technology. Thus, it behooves the education community to develop and deliver music-oriented programming which is educationally and technologically viable, while still being equal to the visual and audio standards of the entertainment products competing for the student's time and interest.

However, we may not have sufficient resources to draw upon to assist us in developing the new technology. In the immediate past, computer-learning research has focused on instructional video or computer software systems from within their role as instructional technology (Clark, 1983; Driscoll, 1984; Heinich, 1984; Petkovich & Tennyson, 1984; Ross & Morrison, 1989) instructional systems design or ISD (Dick, 1977; Dick & Carey, 1985; Gagne, Briggs & Wagner, 1988), and a multitude of interactive learning systems such as CAI (computer-assisted instruction), CBE (computer-based education), CBI (computer-based instruction) and CMI (computer-managed instruction). However, the concentration solely on the form of computer-based delivery ignores the simple fact that the mere presence of a tool is not beneficial if a learner does not understand how the tool can help.

Recent research on learner control, adaptive technologies, and other technical issues still remain hopelessly quagmired in the belief that the instructor will remain somewhat in control of the information processed through these mediums. This false hope does not fit into the reality of the '500 channel universe', where students will very likely pick and visually choose only those subjects which inherently motivate and entertain them in much the same manner as they select their CD's and video games. Nor does it fit into a situation where the instructor is not yet fully aware of the epistemological uses of the tool. Therefore, music educators should begin to develop their understanding as to how the accompanying visual information relates to our perception of music.

In the Western pitch-oriented context, music moves from one note to the next, essentially differentiating one pitch from the next through a series of contrasts, whether they are pitch, duration, dynamic, or otherwise. For example, a middle C moving a tone upward to a D would not only alter from a starting pitch of 261 Hertz to one of 294 Hertz, resulting in a change or contrast in the fundamental frequency as well as harmonic content, but also would probably show some differentiation in the amplitude and duration of the two notes, due to the natural variations in performance. Therefore, it may be reasonable to assume that the modification of the degree of contrast between the notes can make their distinctions more evident or pronounced. The research in Chapter Two seems to bear out this proposition.

However, music appreciation is increasingly becoming a multimedia experience. Many people watch music performances on television, and the music is accompanied, not only by a static camera shot of the performer(s), but by a roving camera, or cameras, each with a different focus or perspective on the performer(s). As well, specific music television channels, such as Much Music in Canada, and MTV in the United States have defined a distinct visual style of music accompaniment; often rapidly-paced, metaphoric, and sometime even comprised of images unrelated to the content or form of the music it accompanies.

Reasons leading to the Study

In the review of the research accompanying this thesis, I found little to suggest that it is a simple matter to successfully combine visual and auditory information - such as a music performance - into a cohesive, entertaining and yet informative manner. Quite the opposite. There appears to be great potential for misperception and wasted opportunity; misperception because the ability to reinforce the wrong information probably increases by - at minimum - a factor of two for every medium we hastily combine together; and wasted opportunity, because we may only get one chance to create something of such significance as an interactive visual / auditory mode of delivery solely dedicated to music education. If we fail to create inspiring and enlightening media, and instead produce those which are ignominiously ignored, our efforts will soon be replaced in the hearts and minds of our students by interactive games and other "educational products" manufactured and broadcast by the likes of Disney or the Children's Workshop. To many educators already accustomed to having many of their favorite composers dismissed as quickly as last year's pop star, the thought of listening to another generation of "Muppet songs" or "Barney tunes" masquerading as educational material is not a very pleasing scenario.

Hence, further research is needed to add to the paucity of current research data attempting to assist us in identifying how visual perception interacts with the auditory in specific musical contexts.

Contributing more information to this data pool could guide broadcasters, Internet programmers, and media developers toward a greater understanding of how to film and/or broadcast education-oriented music performances for the purposes of conveying music lessons, theory instruction, and other forms of music education. This information could help in the decisions on visual camera angles, lighting, presentation methods, pacing, sound and image editing, and other technical issues.

It is believed that research such as this is quite necessary for a number of reasons: the evolving information highway may turn to the type of format used by Much Music, MTV and other commercially oriented television shows for the broadcasting paradigm of music-education TV. MTV, as noted earlier, is mainly recognised for its ultra-short visual edits, with average image durations averaging one to two seconds, as well as its extreme contrasting imagery. The style of MTV has become a paradigm, evolving into a technology of such influence that it has "contributed to the record company reorganisation, increasingly formalising the process of visual composition and generating demand for yet more professional 'mediators' - stylists, publicity and image experts and video directors". (Negus, 1992. p.27) Although it could be argued that this paradigm is not the most effective method to transmit substantive music-oriented education information, it must be recognised that Much Music, MTV, and popular music television generally, has attained a considerable market share; something many "serious" music broadcasts have been unable to do.

These evident successes have led many broadcasters to use the MTV model as a logical stylistic point of departure when they wish to capture the attention of a youthful and impressionable market. Another reason for continuing audiovisual research lies in the very real possibility that governmental funding of education will continue to decline. Therefore, it is reasonable to assume that the necessary funding for research and development of music-education broadcasting and interactivity may come from a partnership with commercial enterprise, which, by its nature, supports only those endeavors which are commercially viable enough to expose them to larger audiences. Larger audiences are usually attained when the product meets or exceeds the audience's stylistic and content-oriented expectations.

It is one thing to base our assessment of the value of the combined image and music on their entertainment value, and yet another on their epistemological value. In order to subscribe an educational value to their relationship, we must determine which purpose we wish them to serve, and then to assess whether that objective has been reached. For example, if the degree of contrast between the musical elements has a significant effect on our perception of the performance, it may also be possible that contrasting the visual accompaniment within itself, and in its relationship with the music it accompanies will also have a significant effect on our perception of the performance.

Thus, further information is needed in order to more fully understand if an MTV-style editing paradigm - or any other paradigm or style of visual editing - creates an acceptable epistemological environment for learning.

For all of these reasons, I wish to undertake a study in order to determine at which speed, at which rate of coincident synchronisation, and at which specific points in time the visual editing of a videotaped performance becomes either beneficial or detrimental to the recognition of pitch errors in the performance.

In order to conduct the study, it was necessary to develop a testable hypothesis. As will be shown in the review of research in Chapter Two, it appears that, regardless of the musical element (intensity, duration, timbre, etc.) - the degree of contrast between it and the other elements in a piece of music exerts a significant influence on our musical perception of not only the contrasted element, but the other elements as well. Contrast appears to focus our attention on a musical element when it is extreme in nature, separating the event from its immediate environment. It appears that the degree of visual contrast creates much the same effect as the degree of sonic contrast. Moreover, the perception of contrast is, in part, based on acculturation - specifically on the role of expectancy and familiarity, and that those most familiar with the context of the contrast will react more acutely to it.

Purpose of the Study

The purpose of this research is to learn about the role of visual editing on our perception of music according to the following principles:

Problems

The specific problems of the study are:

- 1) to determine the comparative effects of four degrees of discrepancy or contrast between the visual and auditory components of a videotaped music performance, and
- 2) to determine to which degree expectancy and familiarity make those most familiar with the context of the contrast react more acutely to it.

Limitations of the Study

- 1) The Teacher and the Investigator are one and the same, which may have produced an adverse effect on the findings.
- 2) The research was conducted using ten volunteer subjects from a single educational institute, and therefore does not constitute a random sampling.
- 3) The study intends only to investigate and catalogue the degree of visual contrast necessary to support or detract from the impact of an education-oriented musical performance.

It is not meant to be a complete study of all the mitigating factors involving audio visual discrepancy, nor involving the recognition of pitch errors, a salient component in the premise of the research. Some aspects involved in music cognition specifically relating to this premise of this study are discussed, including the effect of familiarity, recognition of degrees of contrast within this study's context, and how events initially perceived as familiar can often be misread and misinterpreted when there is sufficient dissimilarity between the visual and audio information.

4) It could be argued that the recognition of a musical element such as pitch errors in no way implies greater understanding or appreciation of the music itself. It is not the author's intention to imply an implicit value to this pitch-related task. Within the circumstance of the study, pitch-errors were, physically, the most simple element to induce.

5) The research found within can only be said to generally apply to situations involving post secondary adult students, and therefore should not be considered generalisable to any other age group.

CHAPTER 2

RELATED RESEARCH

One of the most important functions of perception is to parse the sensory array into objects and to inform the organism about the behavior of these objects. (O'Leary & Rhodes, 1984. p.565)

Thus it is that music is broken down into its composite parts, typically categorised into the elements of harmony, melody, timbre and rhythm, which, in turn, are each comprised of changes or variations in amplitude, frequency, duration, etc..

Related Research on the Perception of Music

Perhaps because of the Platonic notion that specific musical scales have a significant hold over our emotions and actions, there has been great effort given to the categorisation of acoustical/musical elements; however, reducing music to its elements doesn't describe their effect on our perception unless each element is individually analysed as to its impact, and ultimately recombined to examine as a whole. Geringer and Masden (1989) speculated that we respond to musical elements in the following order: rhythm, pitch, tone quality, and loudness. Doane (1989) suggested the following hierarchy (least to most difficult to detect) pitch, dynamics, rhythm, precision.

Byo (1989) suggests a rhythm-pitch hierarchy. Apparently, increases in tempo are not discriminated as accurately as the decreases (Killian, 1985; Masden, 1979; Kuhn, 1974) and tempi tend to increase in performance (Kuhn, 1977; Kuhn & Gates, 1975). Pitch decreases are discriminated more accurately than increases (Geringer & Madsen, 1987), while pitch deviations in performance are perceived as predominantly in the sharp direction (Geringer, 1978; Geringer & Sogin, 1988; Geringer & Witt, 1985), and, regardless of all that, the culture and environment of the listener would still have a significant effect on their musical perception. (Walker, 1987, 1990).

The above studies focus on the prioritising of information, a subject central to this study. Most of them are predicated on their subjects' noticing one of the elements in a certain context. The context - where changes or variations in musical elements such as the alteration in loudness, frequency, etc., are used to highlight structural units, such as phrase boundaries, metrical frameworks, or melodic or harmonic events - centers on whether the element changes within a piece of music, and whether the change conveys something to the listener. However, it must be noted, although expressiveness is typically measured by the intended message getting from the performer to the listener:

In the process of message parsing and recoding, the listener may impose meanings unintended by either the composer or performer. Indeed, it is in the nature of the listener always to seek meaning (Ullman p. 135)

This statement exposes an Achilles heel in elemental music research: our attempts to understand the significance of all or any part of the sensory input is hampered by our inability to discern between what is actually there and what we thought we heard.

Some researchers in musical perception have sought to limit their research to only the most tangible data. Walker (1987) writes that this research has a tendency to uphold some standard as a model against which other data must be measured:

studies in auditory perception are typically concerned with pitch or intensity perception, either involving matching with some given standard or requiring various acts of auditory scaling or categorization. p.491

In this manner, Kendall and Carterette (1990) elected to limit their study of musical perception to the recognition of a single musical element, specifically the recognition of dynamics. In their examination, the subjects were not asked to talk about or write about their perception, instead they were asked to act them out; to express their reaction to stimuli by physical means, which included - upon listening to music - performing the same piece with what they thought were the equivalent dynamics. The subjects (four instrumentalists: oboe, clarinet, violin, trumpet) listened to, and evaluated a concert pianist performing in 3 distinct expressive styles: without expression (mechanical), with appropriate expression, and with exaggerated expression. Each of the four instrumentalists studied a piano reference score of the music prior to the performance.

As they listened to the performance, they were instructed to insert dynamic marking into the score relative to their own impression of the dynamics of the auditioned piece. In the concluding part of the study, the instrumentalists "played the piece in his interpretation of the intended level of expression." (1990, p.137) When the results were measured, all four instrumentalists could only match the pianist's dynamics and timing in the 'appropriate' mode of expression, while no one could match either of the remaining modes, especially the 'exaggerated' mode.

These results lead Kendall and Carterette (1990) to speculate that the instrumentalist's inability to match exaggerated modes may indicate that they needed to perceive a 'natural' timing and dynamic contour in music if they were to accurately understand or imitate it. These natural timings might include a gradual lengthening of successive durations leading to a cadence, and note values which get longer with increasing expressiveness (i.e.: loudness). The answer as to why certain timings and contour are more effective than others remained elusive to the researchers, although they have expressed opinions that a paradigm may lie within us, one which is not based on anything "as strict and invariant as [a] musical grammar, performer grammar, or listener grammar." (p.160) Instead, it could rely on manifold procedures and multiple strategies where "contrast is the key operative principle, not merely accent in terms of increased magnitude." (p.160)

Contrast appears to be an important part in our perception of the elements of music. When Nakamura (1987) investigated the ability of a performer to effectively communicate dynamics to a listener, he found that contrasting dynamics may be easily confused with contrasting pitch. Nakamura's musicians performed a Baroque Trio, and afterward they listened to and notated the dynamic markings of their own interpretation by means of expression symbols (p, f, cres, dim, accel, rit, etc.). These pieces were subsequently listened to by 38 subjects, who were asked to:

choose one of the given possible dynamic symbols or to use their own words, as they considered appropriate, to indicate their own impressions of the performance while listening to it. (Nakamura, 1987. p. 526)

It was found that crescendos are both easier to perform and easier to perceive than an equivalent decrescendo. However, in a number of instances, pitch changes were also perceived as a change in dynamics, such as when a rising pitch was perceived as a crescendo, and falling pitch enhanced the feeling of decrescendo. This pitch-dynamic interrelation led to the confirmation - through acoustic measurements taken with a decibel meter - that there is a significant discrepancy between the musician's intended dynamics and their true performance intensity in dBs. Thus, a performer's intended dynamics rarely relate to the actual measured intensity:

It is to be expected that intensity level is not fixed by a given dynamic symbol, but is influenced by context. It is to be expected also that this effect would appear in the listener's perception. (Nakamura, 1987. p.531)

Therefore, dynamics are not actual levels, but are instead perceptually determined and influenced by changes in context. Additionally, context may also have an effect on the very state of our perception.

Huron (1992) ventures that contrasting loudness may accomplish the perception of crescendo or decrescendo because audiences tend to exist in either of two states of perception: passive attention, where the perceptual event itself captures their attention, and active attention, where attentiveness arises from an self-imposed state of readiness or arousal regarding possible perceptual events.

In the area of passive auditory attention, one factor known to contribute to attention and arousal is changing loudness. (Huron, 1992. p.84)

Could this lead to a conundrum? If changing loudness catches our active attention, while at the same time becoming a status quo because of its continuity (the more things change, the more they remain the same; an old saying goes) how will we know when to switch between passive and active attention? The answer appears to lie in the degree and length of the contrast of the dynamic.

As the stimulus level departs more from the expected or habituated level, the probability of evoking an orienting response increases. (Huron, 1992. p.84)

In contrast to Kendall and Carterette (1990) who found no single element which had a determining influence on our perception of music, Huron (1992) found that "existing research suggests that some patterns of dynamic change are better able to arouse listeners and command attention than other dynamic patterns." (1992, p.36)

This led Huron (1992) to conjecture that a "ramp-archetype" - based on the association between the direction of change of dynamic level and the magnitude of that change - is consciously and/or unconsciously utilised by composers to evoke shifts in their listener's perception. Such musical contrasts are necessary since "as constant escalations of stimulation are impossible to maintain over long periods of time, [thus] an alternative mode of organisation consistent with the maintenance of passive auditory attention is to alternate periods of gradually increasing stimulation with moments of abrupt stimulus decrease." (1992. p.34) Thus, it seems that abruptness of the change or contrast is the key to it having an effect on our perception.

Huron (1990) based his hypothesis on his study of the changes in dynamics found in Beethoven's 32 piano sonatas during which Huron discovered a recurring ramp pattern in Beethoven's dynamics where the music tends to build in a gradual way, but also tends to subside relatively quickly. To further test this theory, Huron (1991) later examined 435 piano works by 14 composers, focusing on their written dynamic markings, the reoccurrence of those markings, their consistency of usage, and their abruptness of change.

From this research it became evident that a number of amplitude-related archetypes were utilised to heighten or maintain interest, such as a clear association between incremental crescendos and abrupt diminuendos. Huron's premise is that acculturated listeners have built a dynamic-oriented archetype into their perception, and have come to expect its occurrence. Consequently, these listeners will react perceptually to those times when the event does not occur.

Although Huron acknowledges that "extant psychological research provides only a vague idea of how large and how frequent these changes would need to be to maintain a listener's attention" (1992, p.86), he cites a study by Mathews (1979) which found that our tolerance to amplitude is effected by its duration. Thus, the perceived strength of the crescendo is directly related to the length of time it is continued to an optimum rate of 10 seconds in duration for maximum effect, after which the effect decreases. Here again, it is not so much absolute contrast as degrees of contrast that assist in the perception of music.

Geringer (1991) supported Huron's (1991) contention that the duration of the event and the degree of contrast between the elements effects our perception of its intensity. In the study, musical excerpts and tones are presented with gradual increases, decreases, or no change in intensity, and subjects were asked to select the category of their discrimination (louder, same, or softer).

Geringer's (1991) research indicated that longer periods of musical silence result in greater difficulty in ascertaining the relative intensity of different passages, and that the entry of a new voice often brings an immediate judgement of increased intensity regardless of any real change in intensity. This seems to support Huron's (1992) finding that the duration of the event changes the degree of attention paid to it.

Thus - if the silence continues too long - it appears to split the perceptual event to the point where the listener feels that they are not listening to a single piece of music with an area of silence within it, but, instead, to two pieces of music bridged by an area of silence. Consequently the listener is unable to maintain the feeling of continuity between the two sections. As well, Huron's noting that the entry of new voices prompts the subjects to make 'louder' judgements sooner than 'softer' responses appears to indicate how contrast leads to increased attention to the event. Here again, Geringer also noticed a number of discrepancies between what the performer hears and what the audience hears with respect to crescendo and decrescendo.

What seems to the performer to be an ample increase in intensity may be barely noticeable to the listener. Correspondingly, a performed decrescendo may be perceived as a much greater modulation than was intended by the performer or composer. (Geringer, 1991. p.119)

While Kendall & Carterette (1990), Nakamura (1987) and Huron (1991,1992) investigated the parameters of louder or softer, Clarke's study (1993) eliminated dynamics altogether in order to ascertain what effect timing and duration had on the perception of musical expression. In a manner similar to earlier research methodologies, a pianist was asked to perform four melodies in 3 different ways; deadpan (without expression), with normal expression, and with exaggerated expression. By disabling the touch-sensitivity on the performer's MIDI device, the performer was prevented from using expressive dynamics in their performance, resulting in musical expression being achieved solely through timing. Noting that imitation tasks have a innate drawback of combining performance tasks with perceptual ones, Clarke wrote:

A primary assumption underlying the use of an imitation task in the experiment reported here is that the subjects attempting the imitations share more or less the same structural representation for the music as the original performer and that when they attempt to imitate his performances, they do so by regenerating the expressive properties from that representation. (Clarke, 1993. p.332)

The contention is that, if expression is generated from a structural representation of the music being performed, disrupting the relationship between the inner timings and duration should make it more difficult for a listener to make sense of the pattern of timing. This, in turn, would make it more difficult for someone to imitate the music, though performance. In order to create yet greater disruption of the original representation, Clarke interpolated MIDI-based transformations of the original performances into the study.

This resulted in melodic inversions, and variations of the inner metric structure, etc., which the subjects were asked to include in their imitation.

The question of whether a failure to imitate correctly is a perceptual or a motor failure reduces to an inquiry about the representation that is common to both processes. (Clarke, 1993. p.318)

In this manner "the transformation that produces the most serious disruption to the relationship between structure and expression in the original performances [such as an inversion of the durations] would also result in the least accurate and most variable imitations." (p.332) Although Clarke ensured that the songs were simple and able to be imitated he nonetheless found little data supporting the need for shared representation between imitator and original performer, a result echoing Kendall and Carterette's (1990) inability to find any kind of shared musical grammar. Clarke instead came to the conclusion that our musical perception is 'ad hoc'; constructed by making up our representation as we go by relating our continual impression with our initial impression. In this manner, as we continue listening, we keep a running commentary on the changes from one point to the next, for example:

A verbal equivalent of this precept for a particular transformation of a melody as represented in a subject's mind might be 'rush the beginning, play the sixteenth notes unevenly, slow up awkwardly in the middle and then rush the end. (Clarke, 1993. p.333)

Other researchers have attempted to determine if we can rely on a perceptual hierarchy upon which to build our continued perception.

Gabrielsson (1985) subjected 30 pieces of music - representing Western art music as well as popular music - to an acoustical analysis, examining various physical properties of the performances including durations, peak amplitudes, envelopes, frequencies, etc. He then created a number of synthesized sound sequences, attempting to systematically vary different factors he observed in the original performances. For example, Gabrielsson (1985) synthesized the theme from Mozart's Piano Sonata in A Major into 3 versions; one being a completely mechanical version where intensities were unchanging, one based on a dynamic profile, another based on a factor analysis of different durations in the 30 performances. Expressive devices, such as the shortening of the longer tone and a lengthening of a shorter tone in a 3/4 or 6/8 Viennese waltz-pattern were employed in the hope of discovering if there was a 'proper' or natural relationship between note durations in certain contexts. These, however, would only prove that:

our perception of such patterns is categorical, at least in part, and that the center and width of the categories are different in different contexts (Gabrielsson, 1985. p.75)

No evidence was found to support an overarching paradigm or hierarchy. Instead, it was reasoned that listener experience or acculturation creates a form of personal categorisation, based on the fulfillment and withholding of expectations.

in overall judgements like these, the ratings of different subjects differ quite a lot, which suggests that their weighing in the experiential dimensions is different. (Gabrielsson, 1985. p. 81)

This, in turn suggests that "there are systematic relationships between performance variable and experiential variables and that the performer makes use of these relationships to effect the listeners' experience." (p.81) Our experience assists us to perceive many of the elements contributing to music, but not all.

Investigating the influence of tone quality and intonation on our perception of music, Masden and Geringer (1981) noted that : the relationship between perceptual discrimination and performance is largely inferential, although aural discrimination ability appears to be a major prerequisite to intonation proficiency. (Madsen & Geringer, 1981. p.305) They began by their producing a tape recording of 24 oboe and flute duet performances, which was subsequently listened to by 480 graduate and undergraduate students with respect to intonation and tone quality. Noting that "it appears that there is greater acuity in detecting pitch flatness combined with a tendency to perform towards greater sharpness" (p.306) the researchers found that nearly all their subjects had the ability to correctly judge the performances that were both in tune and played with good quality with a number of perceptual inconsistencies. For instance, their subjects could not discriminate between what the performers felt to be their best and worst tone qualities within a musical context; an apparent contradiction of earlier studies pointing to the perceptual benefit of contrast within a musical context. (Nakamura, 1987; Kendall and Carterette, 1990; Geringer, 1991; Huron, 1991,1992)

However, further tests revealed that subjects could not perceive intonation errors any better than they could tone quality errors, and that "deviations in intonation as large as 50 cents (.25 tone) between soloists and yet [still] in musical context [were] confused with other considerations" (Madsen & Geringer, 1981. p.312) The researcher felt that a general pattern of misperception emerged whenever the contrasts were not significant enough to be easily or quickly perceived.

As noted earlier by other researchers (Nakamura, 1987; Geringer, 1991), performers often tend to underestimate how much contrast is necessary to induce an audience's perceptual shift. The discrepancy between a performer's and an audience's perception regarding intensity may give rise to the following question: is a musician's perception of music and that of an untrained musician (i.e.: audience member) altogether different? There may be an apparent difference between what the different groups perceive if not how, as shown in Rentz's (1992) examination of 120 subjects' (60 musicians, 60 non-musicians) ability to discriminate between the different textures and timbres of the orchestral families. In the study, the subjects were asked to identify which of 4 orchestral categories of instruments they were listening to during a 6-minute excerpt from Copland's *Billy the Kid*: (a) brass, (b) percussion, (c) strings (d) strings, and (e) all (3 or more orchestral categories simultaneously).

Rentz (1992) determined that musicians and non musicians focused their aural attention similarly with respect to pitch, (supported by Brand and Burnsed's 1991 study indicating that the ability to detect pitch-related music errors may exist independently of other musical abilities); however, the two subject groups responded differently with respect to instrument categories. For example, musicians were found to be able to focus on three or more categories simultaneously, centering on the strings more often than non-musicians. In comparison, non-musicians gave more of their attention to the percussion section, and to the instrument playing the melody. Rentz (1992) writes that this difference may imply that a non-musician has "a lack of sophistication in making immediate aural adaptation to other families following the pronounced use of percussion in the composition". (p.191) These aural adaptations - which sound similar to Clarke's (1993) 'verbal equivalent' concept - would allow the musically-trained - but not the musically untrained - subject to freely move from one aural stimulation to another, without fear of getting confused or lost.

As well, the context of the timbre-stimulus may be a determining factor in the accuracy of its perception. In Byo's study of graduate and undergraduate music majors (1992), subjects were given 3 trials per excerpt to detect 32 potential performance errors (16 rhythmic errors, 16 pitch-oriented) deemed to be 'typical' by public school and university-level instrumental ensemble experts.

For example, it was found that certain voice categories - indicative of frequency and timbre - were more easily perceived while others were perceived with difficulty; for instance, errors within the soprano and tenor voices were difficult for most subjects to detect, while alto and bass-voice pitch errors were more successfully detected. In fact, much of Byo's research (1989,1993) indicates that timbre exerts an influence on the subjects' ability to detect pitch errors, but only within certain contexts, so that:

a pitch error in one context (e.g., polyphonic texture, single timbre, alto voice, melodic function) may have a different degree of detectability than the same error inserted within another context (homophonic texture, multitimbre, soprano voice, harmonic function).
(Byo,1993. p.165)

Thus Byo (1992) noted (as echoed by Gabrielsson, 1992), that context is central to accurately measuring perception; therefore, error detection and perception are best investigated from within a musical framework, and not in laboratory-like isolation.

It is undoubtedly misleading to refer in the example [earlier] mentioned to the 'same error' because, in fact, there are as many different errors as there are contexts within which the error might occur. (Byo, 1993. p.165)

Krumhansl and Iverson (1991) also found that pitch discrimination was enhanced or diminished by the musical context of its presentation, and that our tendency to perceive pitch in relation to other pitches was increased when the pitches preceding or following a fixed target tone were unchanged.

As well, the researchers noted that when the surrounding pitches were held constant while they changed their timbres, their subjects became confused in their memory of the target pitch. In effect, changing the surrounding pitches' timbres diminished or blurs the intensity of their contrast. As mentioned in the beginning of this chapter, reducing music to its elements cannot contribute to our understanding of their combined effect on our perception. Once analysed in isolation, the parts must ultimately be reassembled and given their free rein. The studies cited thus far have pointed out that - regardless of the element (intensity, duration, timbre, etc.) - the degree of contrast between it and the other elements in a piece of music exerts a significant influence on our musical perception of not only the contrasted element, but the other elements as well. Further, it appears that contrast focuses our attention on a musical element only when the isolation or contrast is significant. Yet, if the isolation is of a radical nature, we seem unable to complete the connection between the event and its background. However, this contrast has less effect on those not acculturated to expect it, as well as those less likely to perceive the contrast as less significant than they need to provoke a response.

The effect of this research leads me to believe that if I wished to focus a student's attention on a specific part of a melody, I would be well served to contrast it from the pitches in immediate proximity to it. However, too great a contrast intensity may cause all the other notes to pale in significance, creating a reverse effect, and disconnecting the target note from its surrounding information.

Moreover, the degree of musical contrast that I feel is sufficient may not be enough to my audience's ears. Perhaps this is why we begin to rely on our other senses to assist us in completing the connection.

Related Research on the Visual Perception of Music

It is not surprising, noting the visual bias of modern Western culture, that the psychology of aural perception has been comparatively neglected. (Schafer, p.151)

The phonograph, capable of delivering half the sound/sight combination of a music performance, would eventually combine with the cinema to create the cinema musical, which would, in turn, forever change the way an audience perceived music. Popular performers like Jolson and Caruso, and choreographers like Busby Berkeley gave a relatively unsophisticated audience unfamiliar with operatic extravagances an indication of just how visually exciting music could be. Further impressions - created through the early television broadcasts, such as the classical music broadcasts by Stoykovski and Toscanini, and modern influences such as Elvis films, Beatle movies, Michael Jackson videos, MTV, and today's multimedia musical artists such as Peter Gabriel - have formed an audience who perceive music not just as something you hear, but as something you see and hear. As well, this audience expects to see and hear the experience of music in ways which would have been entirely foreign to all the preceding generations of audiences before them.

In the age of the 'lip-sync'; the age of the televised non-performance, it is quite acceptable - perhaps preferable - for a performer to mimic their music after-the-fact. Perhaps this indicates that the visual content has begun to supersede the musical content in relative importance. Alan Durant (1984) has argued that music has always been experienced as part of the pleasure of seeing, and that music video is a natural return to the enjoyment of seeing and listening. If so, educators should capitalise on the situation by studying the impact that visual imagery, motion and contrast have on our perception of sound.

Earlier, it was suggested that a high degree of elemental musical contrast is necessary in order to focus our audience onto a specific part or pitch within a composition, or song. Yet, this contrast, in its significance, may be antithetical to the composition, might be unmusical, could be simply bad taste. Perhaps we should depart from the idea of imposing some kind of musical contrast on the pitch, and instead look toward imposing some form of visual contrast on it. By using multimedia - where the music and visuals are digitised and eminently editable, perhaps we can point out what we want the audience to hear by pointing out what we want them to see.

Hillstrom and Yantis (1994) found that physical motion can be detected effortlessly when a moving object is sufficiently segregated from or contrasting with its perceptual group, or when the motion appears relevant to the observer's task.

Drawing a distinction between voluntary, goal-directed attention guidance such as when subjects adopt a deliberate state of attentional readiness, and stimulus-driven attention capture, such as when attention is drawn to an attribute that is independent of or irrelevant to the subject's task (see Huron's (1992) passive & active attention in chapter 1), the researchers found that efficient detection of an object or attribute in visual search is the result of instances of goal-directed attention allocation, or the result of extreme segregation or contrast between perceptual objects.

Jonides and Yantis (1988) had earlier found that the detection of a target was markedly enhanced when the target was presented as an abruptly onset character embedded amongst other characters not characterised by abrupt onset; or, in Huron's 1992 terms, a state of active attention sufficiently contrasting with one of passive attention. Hillstrom and Yantis (1994) found that an abrupt appearance of a perceptual object in a previously blank location helps to draw attention to the segregated object, thus:

when motion segregates a perceptual element from a perceptual group, a new perceptual object is created, and this event captures attention. (Hillstrom & Yantis, 1994. p.399)

The contrast exists between the active object and passive background, and is perceived in much the same way a motionless moth remains invisible to its predator until it flaps its wings; the motion helps the predator focus on the prey more effectively.

However, the researchers also found that motion can most effectively guide attention only when it is informative or predictive of the location of a visual search target; it cannot draw attention when it fails to predict the target's position. Therefore, motion must lead the observer to a form of 'visual' conclusion in order to appear useful enough to allocate attention to. This information appears to reinforce the conclusions that motion - whether musical or visual - is most effortlessly perceived when it is in significant contrast to the main body of the work. Additionally, however, motion is not enough by itself; it must also appear to be leading their observer/listener to a useful conclusion about the event causing its action. The concept of useful conclusion is effected by the question 'useful to whom'; to which the answer is 'useful to those experienced enough to develop a hypothesis about the eventual outcome of the unfolding event'.

Thus, it appears a perceptual archetype may be expected, one where we will not watch for or listen to an event when:

- a) it does not go where the we wish it to go, or
- b) it does not do anything in contrast to what we expect it to, or
- c) it does not do anything which appears useful to us.

However, problems with such limited interpretation of the archetype emerge when the interpretation of the event is not accurate, and when we are not really seeing what we think we see, or hearing what we think we heard. And what happens if the two perceptual elements - sight and sound - are interfering with the accuracy of our decision-making processes, and playing a trick on us? It may be that acculturation, or prior experience help form a hypothesis about any emerging action, and this may be needed to separate perceptual fact from fiction. However, research points to a possibility that we do not rely solely upon sound for our perceptual cues, and that visual information may be of equal or greater importance in our perception of sound.

The relationships between the perceptual cues of sight and sound were investigated in the "McGurk effect", a phenomenon observed by McGurk and Macdonald (1978) where it was found that a listener's perception of sound is greatly influenced by sight. They proposed the hypothesis that when viewing a person speaking, the manner of articulation is detected by the ear and the place of articulation is detected by the eye. In their landmark experiments, McGurk and Macdonald (1976) presented normal hearing children and adults with auditory information for one utterance simultaneously with visual information for an alternative utterance. Somewhat unexpectedly, the subjects reported no dominion of one information source over another, but, instead an interactive relationship which produced a third utterance unrelated to the two original information.

For example, given the visual ba-ba and auditory ga-ga, the subjects reported hearing da-da. A later experiment (McGurk and Macdonald, 1978) appeared to indicate that the acoustic signal dominated in absence of the visual, while the visual dominated when the two influences (visual and aural) were present.

In a related study, Delke, Fowler & Funnel (1992) found that whole or real words produced the same result as McGurk and Macdonald's (1976, 1978) experiment with syllables. Delke, Fowler & Funnel's 33 subjects were divided between those who both listened and viewed the spoken word (and were frequently given mismatching visual-auditory cues) and those who only listened to the word without any visual information to assist or confuse their perception. In most cases the amount of confusion was striking:

in the no-view condition, the subjects were very accurate in reporting what they heard the speaker say. In that condition, 97% of the responses chosen were auditory selections, 3% were McGurk effects [signifying a small amount of misperception], and 0% were visual responses. In the view condition, 17% of the responses chosen were auditory selections, 79% were McGurk effects [signifying a great amount of misperception], and 4% were visual responses. (Dekle, Fowler & Funnel, 1992. p.357)

Thus the visual stimuli were capable of causing significant errors in perception. The researchers also found that their subjects were able to adapt and switch between information sources when it became necessary to do so in order to form a construct. The subjects did this by moving their attention to the information source which momentarily appeared most useful or salient.

Delke, Fowler & Funnel found that "visual information for phonetic contrasts is influential for those phonetic properties that are observable visually. Information for some places of articulation are particularly visually apparent; complementarily, acoustic information for place is generally quite labile. When the two information sources conflict, the more reliable source dominates." (p.361)

In a related study by Rosenblum and Fowler (1991) it was found that loudness judgements are effected significantly by visual information even when research subjects are instructed to base their judgements only on what they hear. In their experiment, videotaped hand claps produced at various levels of effort were dubbed so that the optically and acoustically specified effort were generally mismatched. The researchers discovered that "the sight of someone speaking a syllable with a given effort can effect the loudness with which the syllable is heard." (p.981)

Favouring Rosenblum's "ecological theory" (1991) which makes no distinction between speech and nonspeech perception, and "involves the notion that for all types of events (including speech), judgements are constrained by the perceived dynamic properties of the event itself." (p.983) the researchers concluded that the visual event itself - and not prior experience with the event - shapes our perception of it.

Unlike the motor-theory (e.g. Liberman & Mattingly, 1989) relating primarily to speech perception, which states that understanding is formed by extrapolating relevant speech and optic information to form a hypothesis about the intended phonetic gesture of the speaker, the ecological theory relies on no apriori or archetypal knowledge.

Massaro & Cohen (1983) dispute this view, speculating instead that associated prototypes residing in our memory impact on our perception. These prototypes may also bring about our misperception, as when a ventriloquist expects us to watch the movement of the dummy's mouth; bringing us to an erroneous conclusion that movement means action.

Ventriloquists do not throw their voices; rather, a listener's precept is thrown by the visual input of the apparent speaker. (Massaro & Cohen.1983. p.753)

Massaro & Cohen (1983) further speculate that we switch between visual and aural perceptual cues before integrating them into a single perceptual event. Each sense has its own value, in that "auditory information appears to be relevant to all important speech distinctions, whereas visual information is relevant primarily for place of articulation." (p.753) Calling this the FLMP, or Fuzzy Logic Mode of Perception, the research reinforces the concept that action must appear useful, or reminiscent of a previously useful event before it is easily perceived - even if our evaluation proves to be incorrect.

In another study investigating the McGurk effect, Saldana & Rosenblum (1993) tested whether subjects were able to distinguish between the sonic plucks and bows on a cello when given sometimes contrasting or conflicting visual information. After making each judgement, research subjects were instructed to rate the discrepancy between the audio and visual components of the presentation. The data demonstrated that video information had a significant effect on auditory judgements, even when subjects were explicitly told to base their judgements only on what they heard.

In a study of cross-modal organisation, O'Leary and Rhodes (1994) attempted to determine if perceptual systems can utilise converging information from more than one sensory modality to organise the another, mentioned earlier by Massaro & Cohen (1983). Based on the theory that a 'ventriloquism effect' occurs only when and if the position of auditory information effects the localisation of an asynchronous visual stimulus, O'Leary and Rhodes presented their subjects with asynchronous information in the form of high and low dots on a computer screen presented one at-a-time and arranged vertically, and accompanied by asynchronous tones in a manner where highly-placed dots were accompanied by high pitches, and low dots with low pitches. The researchers noted that differing rates of presentation brought about differing degrees of perception, for instance, where visual and sound objects occurred at slower rates, objects were perceived as occurring one-at-a-time.

However, when the rate of presentation was increased in intensity, the upper and lower note and object were perceived simultaneously, or as two-at-a-time. Although O'Leary and Rhodes never withdrew the auditory stimuli in order to determine the influence it played in the perception of seen objects, they nonetheless concluded that "the number of objects seen influences the number of objects heard in an analogous, in phase, covarying auditory sequence. This influence also occurs in the converse direction, with the number of objects heard effecting the number seen." (p.568) These results seem to agree with earlier findings (Mathews, 1979, Geringer,1991; Huron,1991) that the duration of the event changes the degree of attention paid to it, in the way that the listener feels that they are not perceiving a single event if and where significant interruptions occur, but, instead, to two separate events bridged by an intermission. Thus, the subject was unable to perceive any degree of contrast between sections presented at the fast rate, instead perceiving them as a single event.

Earlier, researchers (Jonides and Yantis,1988; Hillstrom and Yantis, 1994) were cited who found that the degree of contrast between visual events significantly influences their perception. Other research (Massaro,1983; Rosenblum,1991)) has referred to the role of expectancy and familiarity in their subjects' responses. These aspects may very well increase our chances of perceiving an event in-and-of-itself, regardless of its relevancy, if only because we know what the event is, and when it will occur, thereby minimising our confusion.

Yet, expectancy and familiarity may have less predictable results; sometimes making us perceive unrelated events when the familiar ones do not occur, such as when Nakamura's (1987) subjects expected a piece of music to fade out, and when it didn't, they got the impression that a crescendo had been produced. In the same vein, Masden & Geringer (1981) have expressed the opinion that "there is the possibility that subjects discriminate quality and intonation differently in familiar melodies." (Madsen & Geringer, 1981. p.311)

In summation, the research in this chapter has indicated that visual perception has a considerable impact on sonic perception, and that contrast is a mitigating factor. It appears that the degree of visual contrast creates much the same effect as the degree of sonic contrast. Moreover, the perception of contrast is, in part, based on acculturation - specifically on the role of expectancy and familiarity, and that those most familiar with the context of the contrast will react more acutely to it. However, even those events initially perceived as familiar can be misread - such as the case of a ventriloquist's dummy - and misinterpreted.

CHAPTER 3

THE EXPERIMENT

Introduction

With the recent advances in teleconferencing, multimedia, and interactive technologies, it is time for music educators and practitioners to look outside the field of music and music education to discover what influences are impacting on our students as they attempt to educate themselves with our help.

With this in mind, I began to research the impact of visual editing on the perception of pitch with the inquiry as to whether or not a musical student would be more likely to focus their attention on a specific point in a melody if they given visually supportive information at that point. In my experience, when I have wanted a student to notice a particular part of a passage of music, I had to:

- a) stop the tape/CD immediately prior to that point, and verbally preparing them, or
- b) provide a musical score, with the passage highlighted, or
- c) relate to theory association, as in: after the 3rd cadence, or at bar 14, etc., or
- d) use musical elements such as dynamics, timbre, or articulation to disassociate the note from its periphery.

However, using these methods as the mode of delivery for musical information is, I believe, not an efficient method if only because the active attention (Huron, 1991) needed for the student to keep track of the data pulls them back from the initial sense of experience; making them observers, instead of experiencers. Perhaps musical contrast, although effective, can be supported by visual contrast, making it possible for the student to be 'led to water' without their even being aware of it.

Subjects

A total of 10 subjects participated, volunteering from the Recording Arts division of Columbia Academy, a multimedia training centre in Vancouver, where they study music engineering, and film and video post-production. Many are experienced musical entertainers and performers, and play one or more musical instruments. They range in age from 19 years to 41 years of age, are equally divided between men and women, and a number of them are immigrants to Canada from Africa and Asia. All subjects have expressed considerable interest in improving their instrumental or performing ability. Almost all have some familiarity with rudimentary music theory. Almost all are more familiar with popular music than they are, say, with jazz, opera, and orchestral music. All have some familiarity with music television such as MTV.

It is believed that the study group closely approximates the skill and motivation levels of a typical "tele" or long distance education student or most students at this level of post-secondary education, but without the instrumental skills, knowledge of advanced musical theory or cohesive music-listening habits of an undergraduate or college-entry music student. A number of the subjects have not completed high school, and, through anecdotal evidence, have no predisposition to ever return to that environment. None of the students had prior post-secondary education. As a private teaching academy, Columbia Academy is not required to set a high academic standard for enrollment, but instead, bases its student selection primarily on evidence of interest. In this way, I believe the student participant in this study further reflects the capabilities of the adult long-distance student, for they, too, may have been unable to fit into a refined academic environment for a number of reasons. Still, they reflect a sincere desire to follow their academically-related goals in their specific region of interest.

Because the intent of the study was to investigate and catalogue the degree of visual contrast necessary to support or detract from the impact of an education-oriented musical performance, it was held that the student's interest in participating in the study made them adequate subjects. Also, the subjects would have little or no pre-familiarity with the early 18th Century Baroque piece prior to the study, making their evaluations of the effectiveness of the manipulated stimuli more like those of an interested, yet uninformed student.

Auditory Stimuli

The 10 volunteers were asked to memorise a specific piece of music for the purposes of the research project. The music used in this study was the 1st movement of Albinoni's Concerto Saint Marc, for trumpet. (for the notation, please refer to the Appendix, Experiment #1 Notation) The material used in the experiment was executed by Martin Berinbaum, a professional trumpet player & Professor of Music at the University of British Columbia. The performance was performed on a Piccolo Trumpet, and was simultaneously recorded on videotape and audio tape at Phoenix Recorders, Vancouver. The trumpet was miked using a Neumann U-47 microphone, distanced 1 meter from, and 6 inches above the instrument. The visual performance was captured to 3 Sony 5850 videotape recorders using a 3-camera system, each camera centered differently on the performer; close-up hands, upper torso with face and instrument, and full torso with face and instrument. The audio performance was recorded on Ampex 467 digital tape to a Sony 3324 tape recorder.

The trumpet was chosen for the research because it is monophonic, unable to play simultaneous pitches, and therefore more easily transposed due to the mechanically-based separation between pitches. It should be noted that none of the subjects play the trumpet. All altering of the original pitches was created by digitally transposing closely-related pitches extrapolated from the original performance using a Macintosh computer running Digidesign Sound Tools, a commercially available digital audio editing software.

Unless used as part of the procedure, none of the actual phrasing was changed; for example, a pitch from the latter part of the performance was not moved into the earlier part of the performance. Instead, the pitches were altered 'where they stood', by manipulating the pitches from within a phrase. The music was digitally transferred back and forth between the Sony 3324 and Sound Tools so that the altered audio would undergo almost no degraded signal-to-noise ratio; information which could be taken by the more sonically astute subjects as a tell-tale sign of alteration.

Procedure of the Experiment

The subjects were given approximately 2 weeks to familiarise themselves with a 2 minute-long cassette audio tape of the melody (for the notated score, see the Appendix). Their instructions were to

"memorise the melody on this cassette. Concentrate on the melody - the rise and fall of the pitch - more than any other aspect of the performance, say the tonal variation, breathing, etc."

They were also expected to be prepared to the extent that they could sing, whistle or perform the melody (their choice) back to the researcher on the day of the research study. On the day of the research study, the subjects were seated in a quiet room approximately 1 meter from a Sony television monitor (see figure 1) and were given a sheet of paper with the following instructions:

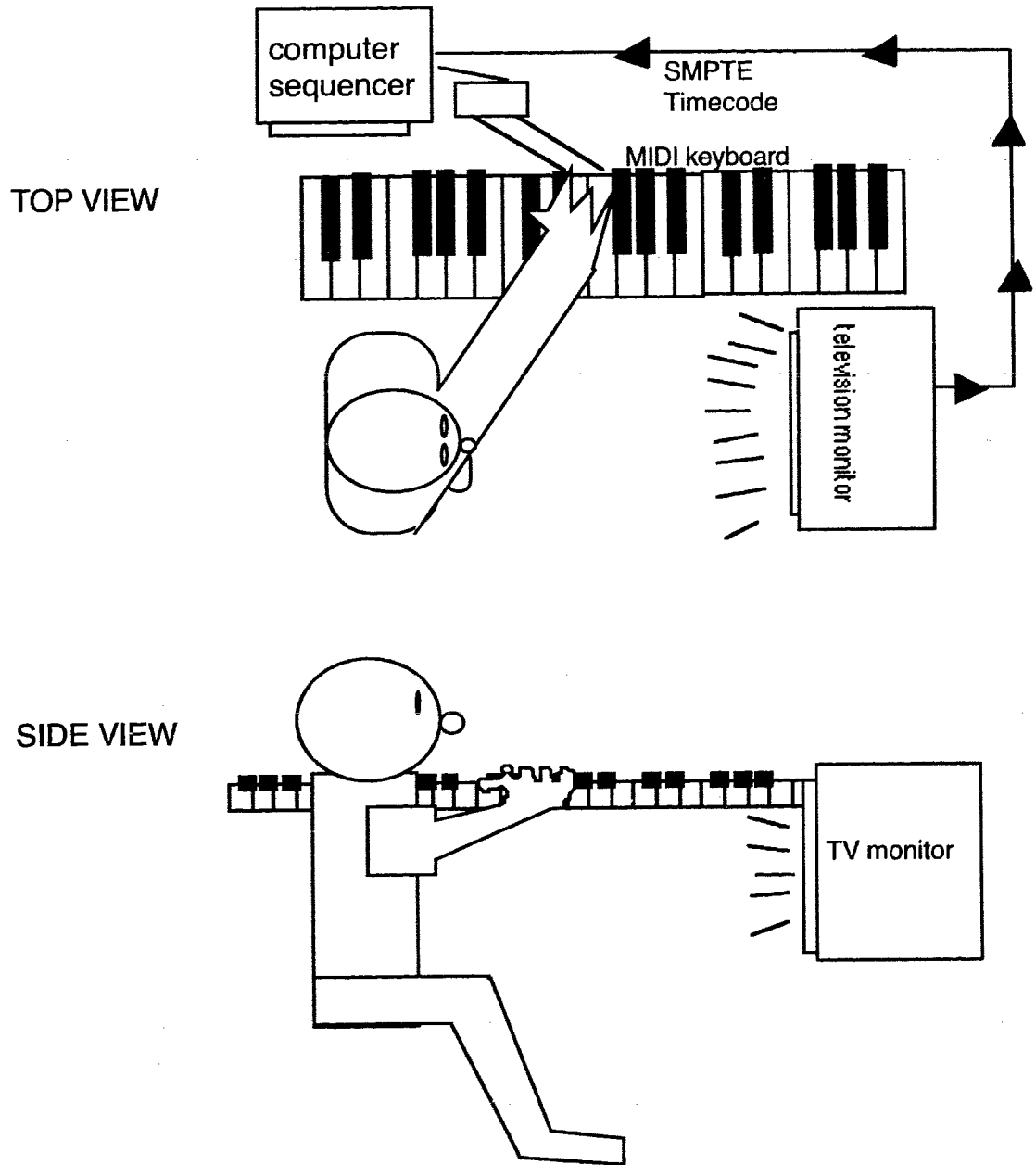
"You have been asked to memorise the melody used in today's research. In a few minutes you will begin watching that same melody being performed 4 times by a musician on videotape. Please make a note of when and if the melody being performed differs from the melody you've previously memorised by tapping on the nearby MIDI keyboard at the time of difference. Although you are to base your judgements solely on what you hear, you should carefully watch the video to note the point of discrepancy."

As shown in Figure 1, the keyboard was linked to a MIDI sequencer and synchronised to the videotape they were watching through SMPTE timecode, noting the exact time, to a 1/2400th of a second, when they perceived the pitch change. Before beginning, any subject who appeared insecure of their knowledge with the melody was asked to aurally demonstrate their familiarity with it before continuing. All the subjects were able to convincingly demonstrate their knowledge of the tune, however subjects #3, 5, 9 and 10 needed to stop periodically during their recitation to refresh their memory. As well, all subjects filled out a brief questionnaire about their musical background and prior musical training.

Based on those questionnaires, 4 out of 10 subjects considered their perception of pitch to be above average, 5 out of 10 subjects considered their perception of timbre to be above average, and 3 out of 10 subjects considered their perception of amplitude to be above average. All the rest of their self-ascribed abilities were average, with the exception of subject #4, who considered her ability to perceive timbre as below average. 9 of the 10 subjects felt their average preparation time of 3 hours was adequate.

Figure 1

Students' seating on the day of Research



After being given a trial run-through, and having acknowledged their understanding of the procedure, the subjects were then presented with a number of audiovisual stimuli, discussed below.

Audiovisual Stimuli

The videotaped music performances were subsequently edited in order to create a number of different visual versions - each version focusing on different points in time relating to the melody. The edited video tapes were then synchronised with the audio tape in order to transfer the audio portion over to the video portion, thus creating an audiovisual performance. The experiment as a whole lasted about 30 minutes for each subject.

There were five audiovisual stimuli in all: one involving no pitch change or induced errors, four involving pitch changes or induced errors. The stimuli was presented sequentially in no specific order. They were compiled into 4 distinct experiments:

Experiment #1: The videotape was edited so that a visual cut was created at the beginning of each new musical phrase. The audio tape of the music was transferred unaltered onto the videotape. The purpose of this tape was to determine if a visual change at the beginning of each new musical phrase might bring about sufficient contrast to confuse the subjects into thinking the melody had changed as well.

Experiment #2: The videotape was edited so that a visual cut was created randomly. The music was transferred altered onto the videotape so that the melody moved away from the anticipated pitch contour (up or down) of the melody in synchronisation with the video edits. The purpose of this tape was to determine if a random visual change could bring about sufficient contrast to assist the subjects in noticing the melody had changed as well.

Experiment #3: The videotape was edited so that a visual cut was created at the beginning of each new phrase. The music was transferred altered onto the videotape so that the 1st note of each new phrase moved away from the anticipated pitch contour (up or down) of the melody by no more than a major third. The purpose of this tape was to determine if a visual change at the beginning of each new musical phrase could bring about sufficient contrast to assist the subjects in noticing the melody had changed as well.

Experiment #4 utilised 2 versions: the 1st five students watched **Version 4a**, similar to Experiment 2, in that the music was edited randomly, but, in this case supported by visual edits discrepant (out-of-synchronization) by as many as 12 frames.

The next five students watched **Version 4b**, similar to Experiment 3, in that the videotape was edited on-phrase, but, in this case the music was supported by random visual edits, in no way synchronous with the audio portion. In both 4a and 4b, the discrepancy exceeded the 80-msec range required to detect an audiovisual asynchronicity.

(McGrath & Summerfield,1985). As well, in both cases, the melody was interpolated with a few additional distractor notes not originally found in the melody. The purpose of both these tapes was to determine if the less supportive video portion detracted from the subject's ability to correctly perceive changes in the audio portion, resulting in less accurate responses, and elongated reaction times.

The premise of the research was loosely based upon the work of Dowling, Lung and Herrbold (1987, 1992), who researched the perceptual effects of listener expectancies with relation to pitch and time. In their 1992 study, experiments were devised to measure the perceptual effects of listener expectancies with relation to pitch and time. In their experiments, trained and untrained musicians (their definition) were presented with a familiar nursery rhyme-like tune such as Frere Jacques through a computer sound generator, and asked to identify the position of interpolated notes, etc., not found in the initial melody.

In their first experiment, a familiar tune was interleaved with 'distractor notes', or notes unrelated to the melody. They found that when the distractor notes were remote in pitch, as opposed to the same relative pitch as the target, it was relatively easy for the listener to pick out the tune. As well, they found that when the 'distractor notes' occurred 'off-beat' or out of synchronisation with the main tune, discernment of the target became more difficult for the subjects.

As mentioned earlier, Krumhansl and Iverson (1991) also found that pitch discrimination of the target note was enhanced by enlarging the contextual contrast of its presentation, in that the tendency to perceive pitch in relation to other pitches was increased when the pitches preceding or following a fixed target tone were unchanged.

In their second experiment, subjects were instructed to notice when a particular pitch (occurring at an expected or predetermined time) moved away from the anticipated pitch contour (up or down) of the melody as the melody was interleaved with distractors. Their results suggested that subjects could focus better on the critical pitch if it moved only a few semitones away from the expected pitch and yet remained within the range of the other notes of the target. However, as the note moved further and further away from the expected pitch, the subjects had greater difficulty in ascertaining where the pitch moved to. As well, subjects had greater difficulty with non-diatonic pitch movement, in which case "a typical observation was that on some of the trials the note seemed simply to disappear." (p.651)

Dowling, Lung and Herrbold felt that their 1992 study indicated that expectation, familiarity and contrast make it more difficult for a listener to judge the pitch of those target notes which fell outside an expected pitch region than those notes which fell within an expected pitch region. Moreover, notes which fell outside the pitch and time windows defined by the expected melody were frequently missed altogether.

Listeners appear to have aimed expectancies in pitch and time at regions where events critical to the identification of melodies are likely to occur [while] the listener's knowledge of a particular tune guides attention to focus on just those times and pitches at which critical events for tune identification should occur. (Dowling, Lung & Herrbold. 1988 p.642)

I felt that their research could indicate that, if our students are familiar with a particular piece of music, they may in fact be somewhat blind to other interpretation possibilities, such as alternative performance dynamics, timbres, pitch options, etc. In this manner, pre-expectation may in fact be synonymous with predestination. Thus, asking a student to anticipate an event to take place, for example, during a music lesson, may in fact be limiting their response to only that specific action. This leads me to believe that we must discover other ways - more subtle, less overt - to focus our students at specific musical points in time; once again, the visual alternative.

The present study may be also considered to be a general inquiry into earlier-cited, audio-related research into the following categories of perception, such as:

Pitch:

Krumhansl and Iverson (1991) noted that pitch discrimination is enhanced or diminished by a note's relationship to those pitches immediately preceding or following it; Dowling, Lung and Herrbold's (1992) found a limited range of pitch perception within a diatonic scale. This led me to include the following into the design:

- altered pitches, both at the beginning of musical phrases and at arbitrary points within the phase, to determine if a melodic cadence preceding the altered pitch made them easier or more difficult to discriminate, and
- altering pitches in a manner that makes it appear as though two concurrent pitches are the same pitch; for example, taking a series of pitches, say e, f and g, and making them g, g, and g through digital transposition.
- limiting the edited pitches to a step-wise distance within a diatonic scale.

Duration:

Geringer (1991) indicated that longer periods of musical silence will result in greater difficulty in ascertaining the relative dynamic intensity of different passages; O'Leary and Rhodes (1984) found that the duration of the event changes the degree of attention paid to it, in the way that the listener feels that they are not perceiving a single event if and where significant interruptions occur, but, instead, two separate events bridged by an intermission. Clarke (1993) contended that disrupting the structural relationship between the inner timings and duration will make it more difficult for a listener to make sense of the pattern of timing; Hillstrom and Yantis (1994) find that physical motion can be detected effortlessly when an object is sufficiently segregated from or contrasting with its perceptual group. This led me to include the following in my design:

- altering (typically increasing) the lengths of silence between musical passages and phrases, to determine if there is a time boundary for misperception or attention given to them, and
- increasing duration at various points within a musical phrase, to determine if there are certain points where our detection is more acute, and
- segregating (visually) certain pitches in a phrase, through differing visual editing, to determine when an object is sufficiently segregated to be effortlessly perceived.

Effect of expectation and familiarity:

Dowling, Lung and Herrbold (1992) found that the listener's knowledge of a particular tune guides attention to focus on just those times and pitches at which critical events for tune identification should occur. Huron (1991) defines two states of perception: passive attention, where the perceptual event itself captures their attention, and active attention, where attentiveness arises from an self-imposed state of readiness or arousal.

Gabrielsson (1991) reasons that listener experience or acculturation creates a form of personal categorisation, based on the fulfillment and withholding of expectations. Clarke (1993) concludes that our musical perception is 'ad hoc', constructed by basing our continual impression on our initial impression. Rosenblum and Fowler (1991) note that the visual event itself - and not prior experience with the event - shapes our perception of it. These conclusions have resulted in this project's being largely based on the research subject's pre or prior familiarity with a chosen melody.

Thus, the subjects are expecting a piece of music to be as it was, and not as it is, in its altered form. Based on these studies, I included the following in my design:

- altering the times at which pitches occur within a phrase, to determine where the critical events for tune identification are, and
- setting up places where the listener might react to the fulfillment and withholding of expectations, by creating micro patterns of alteration, such as continually altering the pitch only at the beginning of each phrase, so that the listener has a chance to react to not only those times when the event does occur, but also to those times when it does not occur, and
- basing the perception of a melody solely on prefamiliarisation, to determine if perception can be undermined, or can erode over time, to determine if the results change the longer the piece of music continues.
- integrating the visual, in order to determine if the visual event alone shapes our perception of it.

Effect of relevancy:

Hillstrom and Yantis (1994) found that physical motion can be more easily detected when the event appears relevant or informative to the observer's task; Delke, Fowler & Funnel (1992) observed that some places of articulation are particularly visually apparent, but when the visual and sonic sources conflict, the more reliable source will dominate the perception.

Massaro (1983) noted that action must appear useful, or reminiscent of a previously useful event before it is easily perceived. Based on this research, I included the following in my research design:

- conflicting, non-synchronous visual and musical information, to determine if physical motion is determined to be more relevant when synchronous, or asynchronous, and
- conflicting visual and musical information, such as when a note occurs sonically but not visually, and vice versa, to determine which is perceived as a more reliable source, and
- setting up micro-patterns which could lead to expectation and familiarity, by creating patterns of alteration, to determine if a visual action must be reminiscent of a previously useful event before it is easily perceived.

Thus my research design included beat-related and random pitch editing; increasing note durations as well as the distance between notes to enlarge note segregation; setting up both the familiarity with and the expectation of musical and visual editing patterns; creating conflicts and contrasts between sound and sight.

CHAPTER 4

RESULTS AND DISCUSSION

Many of the findings of the previously cited studies were confirmed by the results of the present experiment.

The subjects were given a test run-through, asking them to tap the MIDI keyboard (Figure 1) while watching a short piece of edited videotape. Perhaps, because they had all received a number of months of post-production training including the use of SMPTE-synchronised MIDI sequencers, they had no trouble understanding that they were respond to the visual information by tapping the MIDI keyboard. Once their comfort and familiarity with the technology was established, the experiment proceeded.

Experiment #1:

In Experiment #1 a visual edit was created at the beginning of each new musical phrase (please see the Appendix: Experiment #1 Notation) while the music remained unchanged from its original form. The intent of this experiment was to determine whether merely changing the visual component brought about a coincident perception of a change in sound, regardless of whether one has occurred or not. If this were to take place, it may indicate a domination of the visual over the auditory in the subjects' consciousness.

It was anticipated that, because the end of one phrase - in Western melodic and harmonic theory - usually results in the beginning of another, the subjects would expect some form of change at those times, and therefore, be more susceptible to some reaction to a change in the visual stimuli. Because the subjects were asked to tap the keyboard in response to the times when the melody varied from that which they memorised, it was anticipated that there would be clustered responses at the times of the phrase-oriented visual edits. (see Table 1) However, this was not the case.

In total, the subjects erroneously responded 9 times to the 130 visual edits (10 people, times the 13 edits in experiment #1), with no significant difference - determined through an F-Test - between those subjects who initially appeared less familiar with the melody than those who appeared quite familiar.

For future reference, the elapsed time between each of the visual edits was as follows:

Table 1
Elapsed times between edits in Experiment # 1

note	1	2	3	4	5	6	7	8	9	10
sec:	9.76	7.19	7.49	7.46	10.39	16.49	9.26	7.19	4.20	14.46

The edits were relatively evenly paced to allow the subjects a reasonable time window of response. Judging by the fact that their few reactions were randomly placed, the subjects appeared to rely far less on the visual editing and far more on their auditory memory.

Some interesting anecdotal occurrences during the course of the experiment are worth mentioning. Those students whose memory of the melody was not all-encompassing - or in other words, who could not accurately sing the entire melody from start to finish without a little stopping and starting - appeared to have almost no trouble discerning that the melody they learned was unchanged. Whether this is from a sort of reverse-psychological reaction; where it is easier to say 'no, that's not it' than it is to say 'yes, that's it', I could not determine. I talked with those subjects after the experiment, and they said it was quite obvious to them that no sound editing had taken place. The test subjects showed little or no reaction to the visual edits, indicating that they had no difficulty sustaining their original aural recollection of the pre-memorised melody regardless of the contrast of the visual image to their memory.

I can only conjecture as to why the subjects displayed so little reaction. In my observation, this corresponds to a situation whereby a non-musician attending a concert can tell if the performer has made the slightest alteration or deviation from the record they have lovingly come to know and memorise. Perhaps, this may relate to Gestalt memory or rote-learning.

It may also coincide with Gabrielsson's (1991) suggestion that we build our perception based on the fulfillment or withholding of expectation, I could not tell from their feedback because, for the most part, they could not explain why they knew nothing had happened, but only that they knew it was so. Perhaps the major difference between the concert-goer and the subjects is that the music had only recently been given to the subjects to memorise, and the musical style was, intentionally, well outside of their usual listening areas of rock and pop.

The findings in Experiment #1 suggest that visual editing, by itself, may not be enough to cause significant disturbances in a subject's original perception of music. Thus, if through accident or design, we were to edit a videotaped performance on the phrase, purely for aesthetic reasons, it is possible that we will not induce the perception of pitch errors at those points.

Experiment #2:

Experiment #2 followed experiment #1 by approximately 30 seconds. As in experiment #1, there was a visual "count down"- backward from ten - on the screen prior to the actual experiment material, to ensure the subjects were not taken unaware at the onset of the performance.

In experiment #2, the visual and musical edits were created randomly,(see the Appendix: Experiment #2 Notation) at musical points where the digital music transposition could be more easily hidden, such as in places where the notes were somewhat isolated by breathing and phrasing, or on pitches which were more easily transposed due to their timbral qualities, etc. The visual component also mitigated where the sound editing occurred, in the way that Martin's trumpet playing had certain visual cues, such as the relaxation of physical tension, posturing, etc., which allowed for a more natural visual cut. All of the visual edits were placed in synchronisation with the music so that the pitch moved away from the original melody coincident with the video edits.

The intent of this experiment was to determine whether:

1) changing the visual component in synchronisation with induced pitch-errors results in a different response time to the stimuli than, say the unsynchronised editing used in later Experiments #4a and #4b. The faster response time may, in this case, be taken to represent the subject's increased perceptual awareness of a change in sound, or an unconscious response to the change in stimuli.

2) shifting the placement of the edits onto beats other than phrase beginnings and endings resulted in different response time to the stimuli than, say, phrase-oriented editing. If this were to take place, it may indicate that domination of one point in a melody in the subjects' consciousness.

It was anticipated that because phrase beginnings and ending are given special status in Western tonal theory, the subjects would not expect change at purely random points throughout, and therefore would show slower response times to random placing, than the equivalent on-phrase editing found in Experiment #3, and the randomly place, but unrelated visual editing found in experiment #4. Therefore, for the purposes of discussion, the relationship between Experiment #2 and #3 will be discussed in the findings of Experiment #3 , as will that between #2 and #4.

As seen in table 2.1, the group average response times to the pitch 13 alterations in experiment #2 ranged from 2.09 to 3.64 seconds, or a range of 1.55 seconds. However, if edit #4 was to be considered an extreme reaction, the range of all the other responses would narrow to only .64 seconds.

The extreme reaction to edit #4 in Experiment #2 may suggest that the placement of synchronised auditory and visual editing influences a subject's auditory perception. However, whether the musical perception of induced pitch-errors is stronger or more precise and accurate when they occur randomly, or on the phrase, is a subject to be discussed later, in Experiment #3. However, although in experiment #1 visual change was, by itself, not enough to induce the sense of auditory change, Experiment #2 does not conclude whether the combined effect of audiovisual editing on auditory perception produces a more significant response than visual editing.

The group average reaction time for the 13 coincident visual & music edits in experiment #2 was 2.41 seconds, and the group average individual reaction times to each of the 13 edited notes were as follows (in seconds):

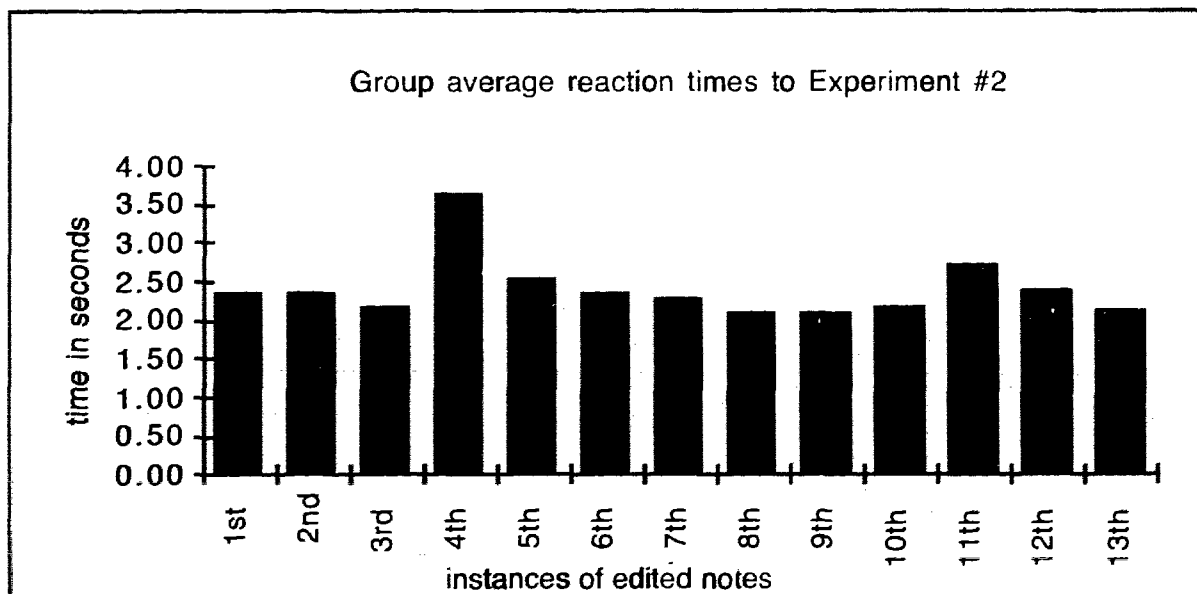
Table 2.1

Group average reaction times to Experiment #2

note	1	2	3	4	5	6	7	8	9	10	11	12	13
sec:	2.34	2.36	2.17	3.64	2.53	2.35	2.28	2.09	2.12	2.17	2.73	2.41	2.15

Figure 2

Group average reaction times to Experiment #2



The time (in seconds) elapsed between each edit was as follows:

Table 2.2
Elapsed times between edits in Experiment #2

note	1	2	3	4	5	6	7	8	9	10	11	12	13
sec:	5.57	7.50	7.86	10.27	1.00	11.67	3.40	6.20	4.97	10.33	7.53	3.30	11.00

Anecdotally, as the experiment transpired, most students reacted comfortably to the now-altered pitches, quickly noting the changed pitches by tapping the keyboard. Perhaps, because they were instructed to base their decisions solely on the music while observing the visual information, and respond at a point of difference between the original and a changed version, the concept no longer took them by surprise. In retrospect, it may have been equally as interesting and informative to have tried the same experiment on a group with general, non specific instruction; perhaps something like "tap the keyboard if something / anything out of the ordinary happens".

As seen in Table 2.1, in visual edit #4 the subjects significantly exceeded their average 2.41 second reaction time. Their increased reaction time of 3.64 seconds may be attributable to the fact that edits 4 and 5 were created virtually back-to-back, with only one second between them after a preceding elapsed time of 10.27 seconds between edits 3 and 4.

Their increased reaction time to edit #4 may lend reinforcement to the findings of a number of previously cited studies. For example, Krumhansl and Iverson (1991) found our ability to perceive pitch in relation to other pitches was enhanced when the pitches prior to and subsequent to a fixed target tone were unchanged. The subjects' slowed reaction appeared to have been due, in part, to the fact that the note closely preceding the target note was pitch-changed. I don't believe that timbre was a factor, due to the fact that the transition was only one semitone, and the digital editing could pitch-shift by that amount with almost no discernible timbral change. Moreover, the music was, intentionally, played back in monophonic format through an inexpensive television speaker, and the subject's discernment of timbral quality would be severely limited.

Relating the findings to earlier-cited research, although the shortness of the elapsed time between edits might be a factor, the elapsed 1 second time (see Table 2.2) between edits 4 and 5 exceeds McGrath & Summerfield's 80-millisecond range required for observers to detect an audiovisual asynchrony. Nonetheless, in practical terms, one second appears to have been too brief, and based on my observations as a teacher, most subjects had difficulty deciphering that two events - and not one - had occurred. At times, they almost looked stunned; knowing that something had taken place and unsure what to do, so they did nothing. Hillstrom and Yantis (1994) found that efficient detection of an object or attribute in visual search is often the result of extreme segregation or contrast between perceptual objects.

It could be argued that one second's time between visual edits would not constitute an extreme lapse of time. Therefore, it may well be that the subjects' inability to detect was partially due to insufficient contrast. O'Leary and Rhodes (1994) noted that differing rates of presentation brought about differing degrees of perception, and that when the rate of stimulus presentation was increased in intensity, two objects were perceived simultaneously, or as one-at-a-time. In this research, as seen in Table 2.2, there were many differing rates of presentation, but none so varying as between edits 4, 5 and 6, (10.27, 1.00, and 11.67 seconds respectively), and this may have contributed to the subject's uneven response to the non-periodic rate of presentation intensity. Milheim (1990) also found that uneven video pacing or pacing outside of the learner's control may lead to dropped motivation. Although I don't feel motivation was effecting their results, I suspect that there is such a thing as visual rhythm, and that certain learners have a propensity for periodic or evenly placed rhythms.

Dowling, Lung and Herrbold (1987, 1992) found that listeners were quite likely to miss altogether any target notes which fell outside an expected time region. As seen in Table 2.2, edits 1,2, 3 and 4 - created at the 5.57, 13.07, 20.93 and 31.20 second marks with elapsed times of 5.57, 7.50 , 7.86 and 10.27 seconds between them - could have been creating an expected time region, leading the listener to an expectation of increasing lengths of time between edits (an increase of 4.70 seconds between edits 1 and 4).

Thus, the one second elapsed time between edit 4 and 5 may have appeared too suddenly, almost to the point of being ignored. Krumhansl and Iverson (1991) found that their subject's tendency to perceive pitch in relation to other pitches was increased when the pitches preceding or following a fixed target tone were unchanged. It could be argued that the (see Table 2.2) 5.57, 7.50 , 7.86 and 10.27 seconds found between the altered pitches 1 through 4 allowed the listener sufficient time to differentiate between the target note and the preceding pitches. As well, only 3 notes occurred in the one second space of time between edit 4 and 5, and the natural pitch in edit #4 was duplicated by edit 5 through digital pitch transposition. This combined lack of contrast and elapsed time may have led to the note being ignored.

As seen in Table 2.1, the second longest average reaction time in experiment #2 was 2.73 seconds for edit 11, an interpolated note not found in the original audio performance. In edit/note #11, Martin Berinbaum's trumpet hand remained visually motionless, yet a change in pitch occurred as the camera stayed on his static hand. Observing firsthand, I noticed that it was those subjects who reportedly had the most or most diverse musical experience who reacted quickly to the implanted note. It could be successfully argued that subjects more familiar with the instrumentation of the pop and rock idiom would be less likely to respond to the act of performing on an instrument associated with the orchestral idiom than they would, say, with a guitar strum or drum beating. Perhaps I could have used a guitarist, playing single notes, or a singer, etc.

However, my initial concern was to create a "level playing field" between the students by distancing them from any music or instrument they had prior associations or personal experiences with. I had hoped that their having two weeks to familiarise themselves with the trumpet solo would, at the very least, have made them begin to mentally associate the sound of the instrument to the method of playing it. Perhaps, because the subjects had been asked to base their judgements solely on what they heard and to note the point of discrepancy, their increased response time could indicate an perceptual conflict of sorts; an uncertainty of what to believe.

The subjects' increased reaction time to edit #11, (see Table 2.1) may lend reinforcement to the findings of a number of previously cited studies. For example Hillstrom and Yantis (1994) found that, although an abrupt appearance of a perceptual object in a previously blank location helps to draw attention to the segregated object (the blank location in this instance, being the previously-blank location now occupied by the interpolated note) such motion will most effectively guide attention only when it is informative or predictive. In the case of edit #11, there was a significant conflict between the visual and audio which was neither informative nor predictable, leading to indecision and increased response time. Delke, Fowler & Funnel (1992) found that their subjects were able to adapt and switch between information sources when it became necessary to do so in order to form a construct. Perhaps it was unclear which source was the most informative, thus, the switching between sources may account for the increased reaction time.

However, there may be a conflict with at least one previously cited study. Macdonald and McGurk (1988) found that the acoustic signal dominated in absence of the visual, while the optical dominated when the two influences (visual and audible) were present. It could be argued that, if the optical dominated, the subjects would not have responded to the visual / audio discrepancy, ignoring the interpolated audio pitch by focusing on the fact that Martin's trumpet hand never moved. Instead, it appears that the audio dominated their perception to a great enough degree for them to slow down long enough to notice the presence of a new pitch. However, because Macdonald and McGurk's research was a seminal project into speech perception, and not intended to relate to musical pitch discrepancy, the conflict may be understandable, and perhaps further clarified by Krumhansl and Iverson's (1991) later research into contextual contrast.

Experiment #3:

In Experiment #3 a visual edit was created at the beginning of each new musical phrase (please see the Appendix: Experiment #3 Notation) In experiment #3, visual edits were created at the beginnings or ends of each musical phrase, points where Martin physically relaxed, breathed, or shifted his physical posture to accommodate a new hand position. The visual edits were synchronous with the music, which had been digitally transposed and transferred onto the videotape so that the pitch-shifting was in sync with the video edits. This method of editing was meant to create extremely supportive audio-visual information.

The intent of this experiment was to determine whether:

1) changing the visual component in synchronisation with induced pitch-errors results in a different response time to the stimuli than, say the unsynchronised editing used in Experiment #4a or #4b. The faster response time may in this case, be taken to represent the subject's increased perceptual awareness of a change in sound, or an unconscious response to the change in stimuli.

2) shifting the placement of the edits onto phrase beginnings and endings resulted in different response time to the stimuli than, say, randomly placed editing. If this were to take place, it may indicate that domination of one point in a melody in the subjects' consciousness.

As in Experiment #1, it was anticipated that, because of the special status given to phrase beginnings in Western tonal theory, the subjects were conditioned to expect change at those points.

Therefore, they would exhibit faster response times to on-phrase editing than those that were randomly placed, such as in Experiment #2.

The findings in Experiment #3 suggest that, as in experiment #2, the even group average response times to the stimuli (see Table 3.1) indicate that synchronising the auditory and the visual planes enhances the accurate perception of music.

Thus, editing a videotaped performance on-the-phrase may be a productive manner to draw their attention to a particular point in the performance.

The subjects in experiment #3 had a more consistent response time than in Experiment #2. Not only was their group average response time faster than experiment #3- averaging 2.06 seconds, or .35 seconds faster than experiment #3's 2.41 seconds - but it also had a considerably smaller range of only .2 seconds (from a 2.12 largest and 1.92 smallest; see Table 3.1) versus experiment #2's 1.55 second range (from a 3.64 largest and 2.09 smallest; see Table 2.1) .

In part, the subject's relatively smooth group reaction time could be attributed to the fact that the 10 edits were evenly dispersed, with time gaps no smaller than 4.2 seconds between them (see Table 3.2), thus minimising the effect of surprise. It may also be that placing the edits on the phrase assisting the subjects in more easily locating where the music differed from that which they memorised.

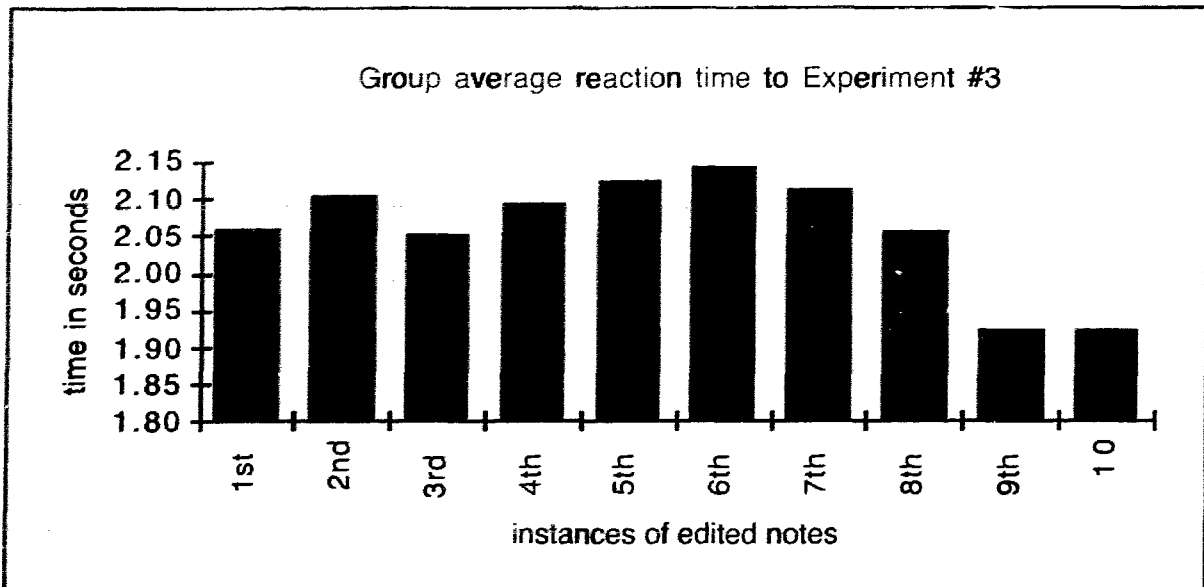
The group average reaction time for the 10 coincident visual & music edits in experiment #3 was 2.06 seconds, and the group average reaction times to each the 10 edited notes were as follows:

Table 3.1
Group average reaction times to Experiment #3

note	1	2	3	4	5	6	7	8	9	10
sec:	2.06	2.11	2.05	2.09	2.12	2.14	2.11	2.05	1.92	1.92

Figure 3

Group average reaction times to Experiment #3



The time (in seconds) elapsed between each edit was as follows:

Table 3.2

Elapsed times between edits in Experiment #3

note	1	2	3	4	5	6	7	8	9	10
sec:	9.00	8.37	8.46	11.40	6.60	16.53	9.53	7.17	4.20	14.10

Anecdotally, as experiment #3 transpired, most students reacted more comfortably to this project than to those preceding it. It is possible that, as seen in Table 3.1, the first 3 edits were 9.00, 8.37 and 8.46 seconds apart. The long periods of time between the edits may have set a pattern of familiarity and expectation, resulting in the even response times to the experiment. Perhaps the only notable feature was that the longest response time of 2.14 seconds corresponded with the longest distance between edits, 16.53 seconds. Although this occurrence may be coincidental, it may also reinforce a number of previously cited studies, for example, Geringer (1991) found that longer periods of musical silence result in greater difficulty in ascertaining the relative intensity of different passages. Although silence may not be equitable with inactivity (as per visual edits), it is possible that the elongated time preceding edit #6 broke with the relative consistency of previous edits, thereby splitting the continuity of the perceptual event. In this same respect, O'Leary and Rhodes (1994) found that longer durations between events lead to the perception of two separate events bridged by an intermission. The extended edit time preceding edit #6 (see Table 3.2) may have led to a slightly less accurate perception. As well, Clarke (1993) concluded that we make up our representation as we go by relating our continual impression with our initial impression. In this case, the initial impression was built up by considerably shorter and more regular edit times of edits 1, 2, 3 and 4 than that preceding edit #6.

Comparing the reactions to Experiments #2 & #3

However different the subject's responses to experiment #2 and #3 may appear, when comparing Experiment #2 to #3 in both an F-Test: Two-Sample for Variances and a T-test, it found that there was no significant difference in their reaction to the stimuli. In 9 out of 10 subjects, an F-test ($p < .05$) showed a non significant difference of 0.00. Thus there appears to be little difference between any one subject's response time to a random, versus an on-phrase editing of a visual performance.

Experiment #4: introduction

In experiment #4, the research group of ten subjects was split into 2 groups of 5 subjects. The first five students watched Version 4a, similar to Experiment 2, in that the music was edited randomly and was supported by visual edits out-of-synchronisation by as many as 12 frames. The next five students watched Version 4b, similar to Experiment 3, in that the videotape was edited on-phrase, but, in this case the music was supported by random visual edits, in no way synchronous with the audio portion. As well, the digitally-edited melody was interpolated with a few additional distractor notes not originally found in the melody.

The intent of experiment #4a and 4b was to determine whether:

1) changing the degree to which the visual component is in synchronisation with induced pitch-errors. This should result in a different response time to the stimuli than, say the synchronised phrase-oriented editing used in Experiment #1 and 3, or the synchronised random editing used in Experiment #2. A different response time may be taken to represent a change in the subject's perceptual awareness of a change in stimuli.

2) shifting the placement of the edits away from any direct relation to the melody results in different response time to the stimuli than, say, melody-related editing. If this were to take place, it may indicate the need for a relationship between the melody and its visual accompaniment in the subjects' consciousness.

The method of editing in experiment 4a was meant to create non-supportive audio-visual information in the way that the visual editing was significantly out of synchronisation with the audio.

Experiment 4a

In Experiment 4a the group average reaction time for the 13 non-coincident visual & music edits was slower than any before - averaging 2.81 seconds, or .75 seconds slower than experiment #3's 2.06 seconds (see table 3.1), and .40 seconds slower than experiment #2's 2.41 seconds (see table 2.1) The range of reaction was also rather wide: 1.8 seconds (4.03 second longest and 2.23 shortest) compared to Experiment #3's smaller range of only .2 seconds (see Table 3.1) although very close to the similarly randomly-placed edit in experiment #2's 1.55 second range (see Table 2.1) .

The slower and irregular group reaction time could be attributed to the fact that the 10 visual edits, although evenly dispersed, appeared somewhat unrelated to the music due to the imprecision of the synchronisation. It could also be that placing the music edits randomly without giving them closely synchronised visual editing caused yet further confusion in identifying the placement of the edits.

The group average reaction times to the 13 edited notes in Experiment 4a were as follows:

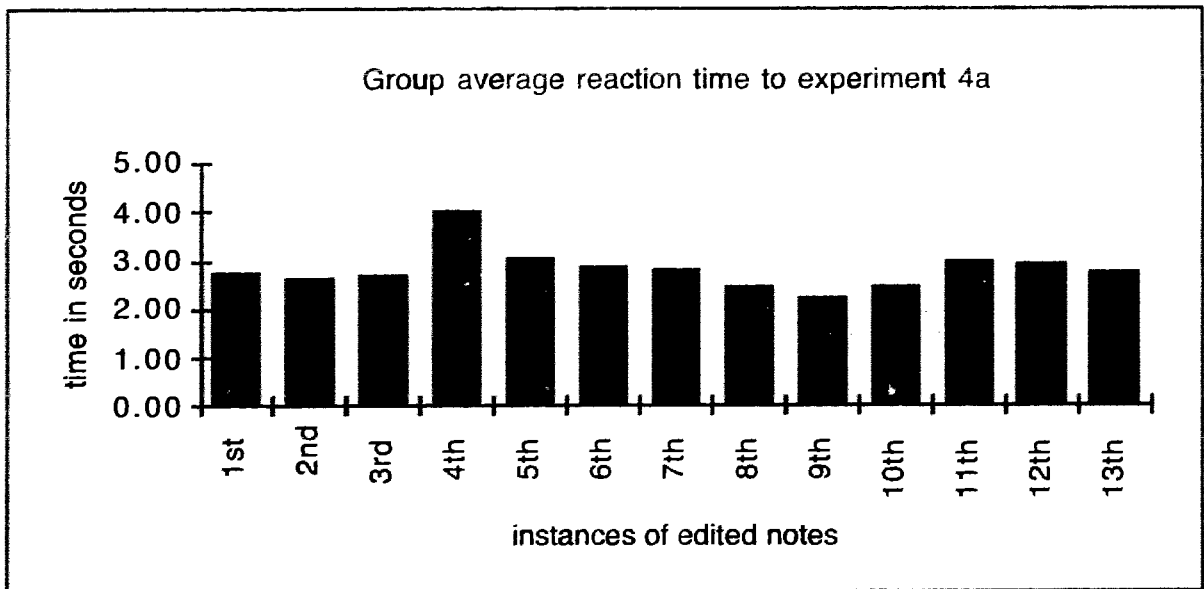
Table 4.1

Group average reaction times to Experiment #4a

note	1	2	3	4	5	6	7	8	9	10	11	12	13
sec:	2.77	2.65	2.66	4.03	3.05	2.87	2.81	2.45	2.23	2.45	2.97	2.89	2.74

Figure 4

Group average reaction times to Experiment #4a



The time (in seconds) elapsed between each edit was as follows:

Table 4.2

Elapsed times between edits in Experiment #4a

note	1	2	3	4	5	6	7	8	9	10	11	12	13
sec:	5.56	7.47	7.70	10.2	1.54	11.4	3.36	12.2	9.70	10.4	6.62	3.92	8.01

Anecdotally, as experiment #4 transpired, most students appeared to react less comfortably to the unsynchronised editing. As seen in Table 4.1, in 5 of the 13 visual edits, numbers #4, 5, 6, 11 and #12, the subjects exceeded their average 2.81 second reaction time. Of interest is the fact that none of the ten subjects reacted to the 4th visual edit. As in experiment #2 (see Table 2.1), the subject's lack of response to edit #4 could be attributable to the close proximity of edits #4 and #5.

It is also possible that the preceding elapsed time of (see Table 4.1) 10.20 seconds between edits 3 and 4, both of which aspects were discussed in some length as a similar part in experiments 2 and 3. As such, the extended reaction times in edits #4, 5, and 6, may be attributed to the radically different rates of presentation (10.20, 1.54, and 11.40 seconds respectively).

Comparing the reactions to Experiments #2 & #4a

A comparison between the reaction times of similarly-based experiments #2 (edited randomly) and #4a (edited randomly, and out-of-sync) shows an average increased response time of as follows:

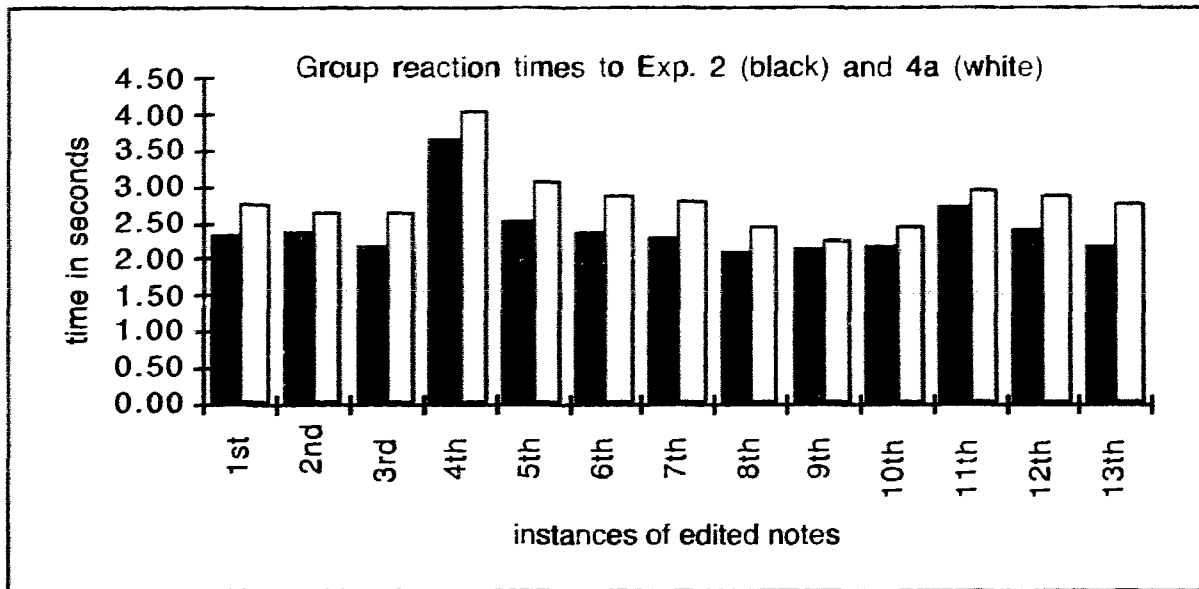
Table 5

Group average reaction times to Experiment #2 vs. #4a

note	1	2	3	4	5	6	7	8	9	10	11	12	13
ex.2	2.34	2.36	2.17	3.64	2.53	2.35	2.28	2.09	2.12	2.17	2.73	2.41	2.15
ex4a	2.77	2.65	2.66	4.03	3.05	2.87	2.81	2.45	2.23	2.45	2.97	2.89	2.74

Figure 5

Group average reaction times to Experiment #2 vs. #4a



When subjecting experiment #2 and #4a to an F-test, the following results indicate a significant difference between the subjects' response times.

Table 6

F-test: two-Sample for Variances; Experiment #2 vs. 4a

STUDENT 1:	= 0.42, or	= p >.05
STUDENT 2:	= 0.37, or	= p >.05
STUDENT 3:	= 0.36, or	= p >.05
STUDENT 4:	= 0.15, or	= p >.05
STUDENT 5:	= 0.13, or	= p >.05

As seen in Table 6, 5 out of 5 subjects appeared to need more time to react as the pitches moved further from the beginnings of a phrase, and were made more optically displaced, distant, and out-of-sync. Based on the data in Table 5, it is reasonable to assume that pitches which were randomly altered are also made yet more difficult to hear by also being visually out-of-sync.

Experiment 4b

In Experiment 4b the group average reaction time for the 13 non-coincident visual & music edits was the slowest yet, averaging 2.90 seconds, or .09 seconds slower than experiment #4a's 2.81 seconds, (see table 4.1), .84 seconds slower than experiment #3's 2.06 seconds (see table 3.1), and .49 seconds slower than experiment #2's 2.41 seconds (see table 2.1)

The average combined group reaction time for the 10 coincident visual & music edits in experiment #4b were as follows:

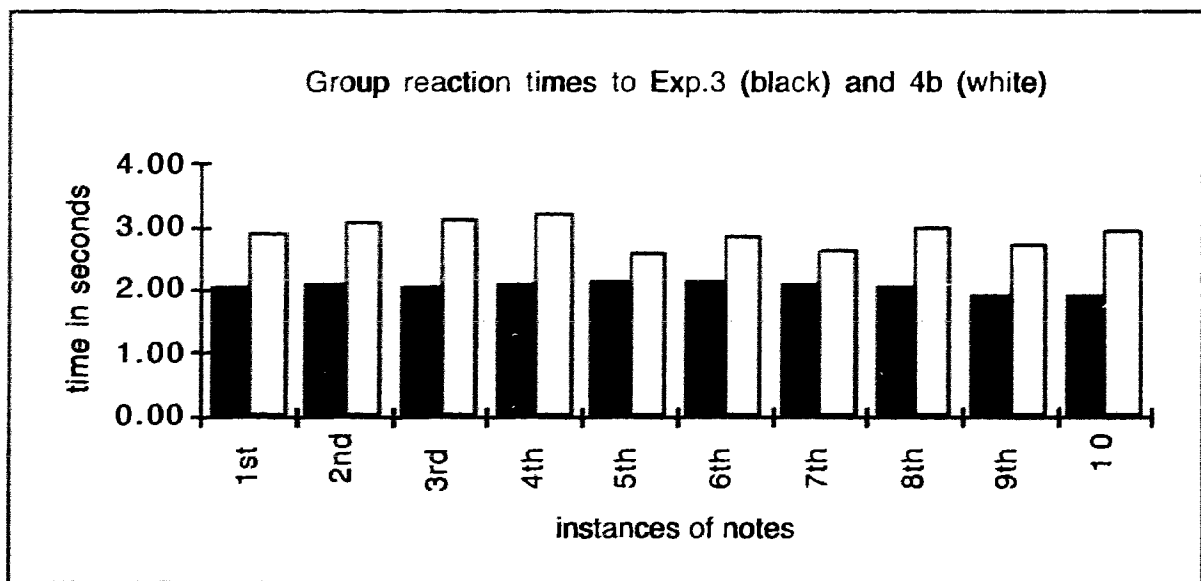
Table 7.1

Group average reaction times to Experiment #4b

note	1	2	3	4	5	6	7	8	9	10
sec.	2.90	3.07	3.10	3.19	2.60	2.86	2.63	3.00	2.71	2.95

Figure 6.2

Group average reaction times to Experiment #3 vs. #4b



In Table 7.1, the subjects show markedly longer average reaction times to audio-visual data which is not closely synchronised, when compared to those closely synchronised. However, it should be noted that, in experiment #4b, as in experiment #3, the range of reaction was also rather small:

59 seconds (3.19 longest and 2.60 shortest), which is only .39 seconds wider than experiment #3. When subjecting experiment #3 and #4b to both an F-test and T-test, the results indicate no significant difference between the subject's average response times. This seems rather surprising, given that experiment #3 not only had perfectly accurate synchronisation, but also on-the-phrase editing as well. I would have expected a considerable more diverse response to the erratic audiovisual material presented in experiment #4b. Judging by this comparison, we may be equally predictable in our response to audiovisual information which is either:

- a) completely supporting, in the manner that it is tightly synchronised, and on the phrase, or
- b) completely unsupported, in the manner that it is neither synchronised, nor related to any musically important points.

Although perhaps not as clearly as in experiments #2 and #3, the significantly increased reaction time to experiment #4b might support a number of previously cited studies. For example, Huron (1991, 1992) found that the listeners will react perceptually to those times when the event does not occur. Perhaps, in this case, the response to an event not happening is similar to that when it does. Gabrielsson (1988) found that listener experience or acculturation creates a form of personal categorisation, based on the fulfillment and withholding of expectations. The subjects may have built up an "acculturation of synchronisation", much as the public has with respect to film, and television.

Most people have little tolerance for an out-of-sync film, and the research subjects' confusion and frustration may have led them adapt a laissez faire attitude, bringing about the same relaxed responses to #4b as to the well supported and cogent #3.

However, I did note more physical signs of concern (sighing, swearing, shrugging, and so on) during the 4th experiments than during the preceding experiments, as well as more inquiries of how "well" they did afterward.

Conceivably, in the manner that Jonides and Yantis (1988) found that the detection of a target was markedly enhanced when the target appeared abruptly from within a comparatively stable environment, no difference in the abruptness of the onsets of the notes due to their lack of synchronisation means that detection increases as the abruptness decreases. Perhaps, as Massaro (1983) speculated, because we switch between visual and aural perceptual cues before integrating them into a single perceptual event, a significant conflict between the visual and aural makes it impossible to integrate them, thus nullifying any confusion.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

Telephone lines should connect every classroom to the Internet and other electronic services. Students would then gain access to resources few schools can afford; they could then communicate with students and experts around the world. It also will support the teachers - perhaps the only group of professionals in our society expected to do their job with no office, no telephone, no privacy. Telephone lines in the classroom would enable teachers to use the telephone like other professionals do: to access information and communicate with community members, peers, parents, students and experts. This is the first step, however. We have talked with student, parents, educators, business, community and political leaders to formulate a vision for education in the 21st century. This vision stresses lifelong learning and is student-centered, making teachers, families and communities co-facilitators of learning. This vision takes advantage of new technologies to support the educational process. (George Lucas, Wired magazine, 1994)

Lucas' vision appeared to be quite distant in 1994, but now, almost two years later, the idea seems rather quaint, and to most people, reasonable and achievable. The explosion of the Internet into our everyday lives, and the overwhelming amount of information that it provides us with makes it more understandable and desirable that multimedia-oriented education be a part of our present, not our future. Just as teachers both past and present supplement their curriculum using videotape, so will both present and future teachers use multimedia to delivering the same information in a multidimensional, non-linear, non-hierarchical, fluid and interactive form.

This study suggests that there is much to research, in terms of how we perceive audio and visual information in an educational setting, and how we can use multimedia in our everyday lives as teachers. On the basis of the experimental data acquired from this study, the following conclusions may be drawn:

Conclusions

1) We may be able to edit a videotaped performance at the beginning or ending of musical phrases, purely for aesthetic reasons. This conclusion is based, in part, on the results of Experiment #1, where visual edits on the phrase, by themselves, did not make the students perceive pitch changes where there were none. Thus it would appear that placing them there is incidental to perception.

2) Synchronising both the visual and audio information so that it occurs at the same time may increase the response time of the viewer to the information at those points. This conclusion is based, in part, on the results of Experiment #2, where the evenness of the subjects' group average response times to the stimuli (see Table 2.1) indicate that it is not merely the placements of the visual edits that assists in their location, but the combined effect of both the auditory and the visual.

3) Synchronising the visual and audio information so that they occurs at the same time, as well as placing the edits at the beginnings and endings of musical phrases may further increase the response time of the viewer to the information at those points. This conclusion is based, in part, on the results of Experiment #3, where the subjects in experiment #3 showed a more consistent response time, as well as a smaller range of response times than in Experiment #2.

4) Dispersing the audiovisual edits in an even manner may shorten the response time to whatever information is at those points. This conclusion is based, in part, on the results of edit #4 in experiments #2 and #4. In both of these experiments edit #4 was overlooked when it appeared suddenly after a number of evenly dispersed edits. However, in Experiment #3, when each of the edits was evenly dispersed, with time gaps no smaller than 4.2 seconds between them (see Table 3.2), the response time was shorter than that of experiment #2.

5) Any form of auditory and visual asynchronicity may bring about slower response times to the audio information. This conclusion is based, in part, on the results of Experiment #4a and 4b (see Table 6 and 7.2) where 5 out of 5 subjects needed more time to react as the altered pitches moved further from the beginnings of a phrase, and were made more optically displaced, distant, and out-of-synchronisation.

6) Audiovisual confusion may have a distinct duration, regardless of its cause. This conclusion is based, in part, on the results of Experiment #4b, where experiment #3 and #4b were both subjected to an F-test and a T-test, showing no significant difference between the subject's average response times in either case. Such a response, or lack thereof, may indicate that, without supporting audiovisual information we respond arbitrarily in kind, yet predictable in degree.

7) It appears as though there may be limits to how quickly we can process combined visual and sound information. In experiment #1, the sound and visual information - although discontinuous in the way that the visuals changed while the music did not - never occurred at a rate faster than 4 seconds apart. However, in experiments 2, 3 and 4, the faster paced editing between edit #4 and #5 (approximately one second apart) appeared to confuse and slow the subjects' response down. It should be remembered that, in experiment #1, the subjects had little trouble retaining their original memory of the tune, and even in the three other experiments, they were generally able to respond at those places where the music was edited, except for those places where the edits were at rates approaching one per second. Conclusion #7 has some unsettling implications, especially if we take into account that MTV-style visual editing is very fast-paced. If our intention is to transmit visual information which supports the music it coincides with, we may, in fact, be decreasing the cognition and understanding of our audience due to the rapidity of the MTV-style delivery method.

This study may have indicated that our intended audience - in an educational setting - cannot accurately ingest combined sound and visual information at certain rates, particularly if modeled after an MTV format. As noted in Grimes' research (1990), although we are capable of achieving faster recognition of events when they are displayed in tightly synchronised situations - even as fast as .5 of a second - retention rates are still poor.

Generally, it appears that we rely on the degree of contrast between the elements - both visual and musical - to assist us in separating a target event from its immediate environment. However, if the isolation is of a radical nature, we seem unable to complete the connection between the target event and its background, as shown by the lack of reaction to the phrase-oriented visual editing in experiment #1, and the extended response times to experiments #4a and #4b. Such a form of contrast appears to have less effect on those not expecting it to take place, and on those whose familiarity with its circumstance makes them less likely to perceive the contrast as significant enough to provoke a response.

Recommendations for further Research

- 1) Although not researched in this study, I believe there is a need to investigate the manner in which, say, North American students respond to multimedia compared with those students from other countries.

When Beenjtes (1988) attempted to replicate Solomon's American-based media study, he found that, unlike their American peers, Dutch children do not unconditionally perceive television as an easy, or "superficial" medium, but, instead have to invest some effort into understanding it. Beenjtes found that one of the most important distinctions between the Dutch and American children, is that the Americans perceive themselves to be much more capable of learning all or most of the topics [they need to know] from television than from books, and act accordingly, in the manner that they 'channel hop', perform other tasks while watching, and indulge in frequent breaks. Whereas, in comparison, because Dutch children do not believe this to be so, they invest more sustained effort in their consumption.

I mention this study to draw attention to the fact that, although the Internet exists on a multi-cultural and multi-national framework and provides a highly democratic base for educators to work with, we may risk adopting the overarching influence of American media and multimedia archetypes unless we research and develop our own.

Walker, (1987) in a study concerned with auditory stimulus differentiation among four acoustic parameters (frequency, waveform, amplitude and duration) found that culture, environment and musical training play a significant role in the consistency of choice for metaphors emphasizing the vertical placement of frequency, certain patterns for waveforms, the size of amplitude, and the horizontal length of duration.

Thus, a central issue as we develop from a North American perspective, is how our intended audience - students from within the Canadian multicultural community - respond to those metaphors and iconographs we choose to represent music.

2) It would be informative if polyphonic, or chord-based instruments could be researched, to determine whether we perceive chords shapes on both familiar and unfamiliar instruments which included the variables of pitch, amplitude and duration. The research could also be broadened to include the depiction and support of musical metaphors, visual metaphors, and non-literal interpretations of both musical and visual information in various combinations.

3) It is essential that much of the research into the effects of contrast, context and familiarity on the visual perception of music be performed outside of the lab, by teachers and students using information technology in the spontaneity and immediacy of their everyday lives.

Bibliography

Augustine. (1964) In Hofstader, A. & Kuhns, R. ed. *Philosophies of Art and Beauty*. The University of Chicago Press.

Brand, M. & Burnsed, V. (1981) Music Abilities and Experiences as Predictors of Error-Detection Skill. *Journal of Research in Music Education*, 29, 91-96

Beentjies, J. Learning from Television and Books. A Dutch replication study Based on Solomon's Model. *Educational Technology, Research & Development*. 37, 46-58

Byo, J. L. (1993) The influence of Textural and Timbral Factors on the ability of Music Majors to Detect Performance Errors. *Journal of Research in Music Education*, 41, 156-167

Byo, J. L. (1992) Recognition of Intensity Contrasts in the Gestures of Beginning Conductors. *Journal of Research in Music Education*. 38, 157-163

Clarke, E.F. (1993) Imitating and Evaluating Real and Transformed Musical Performances. *Music Perception* 10, 317-341

Clarke, R.E. (1983) Reconsidering research on learning from media. *Review of Educational Research*. 53, 445-459

- Dekle, D.A., Fowler, C.A. & Funnel, M.G. (1992) Audiovisual integration in perception of real words. *Perception and psychophysics*. **51**, 355-362.
- Dick, W., (1977) *Formative Evaluation*. In L.J. Briggs (ed) *Instructional Design: Principles and applications*. Englewood Cliff, Educational Technology Publications.
- Dick, W. & Carey, L., (1985) *The systematic design of Instruction*. Glenview, Il. Scott, Foresman.
- Dowling, J.W. Lung, K, & Herrbold, S. (1987) Aiming attention in pitch and time in the perception of interleaved melodies. *Perception and psychophysics*. **41**, 642-656.
- Drake, A.H. (1968) An Experimental study of selected variables in the performance of musical durational notation. *Journal of Research in Music Education*, **16**, 329-338
- Driscoll, M.P. (1984) Alternative paradigms for research in Instructional Systems. *Journal of Instructional Development*. **4**, 189-194
- Gabrielsson, A. (1985) Interplay Between Analysis and Synthesis in Studies of Music Performance and Music Experience. *Music Perception* **3**, 59-86

Gagne, R.M., Briggs, L.J. & Wagner, W.W. (1988) *Principles of Instructional Design*. New York, Holt, Rinehart Wilson.

Gardner, H. (1988) *Toward More Effective Arts Education*. In *Aesthetics and Arts Education*. eds. Smith, R and Simpson, A. University of Illinois Press.

Gardner, H. (1986) *The Anthropology of Music*. New York, Norton and Norton.

Geringer, J.M. (1978) Intonation performance and perception of ascending scalar patterns. *Journal of Research in Music Education*. **26**, 32-40.

Geringer, J.M , & Masden. C. (1989) Pitch and Tone quality discrimination and preference. *Canadian Music Educator*, **30**, 29-38.

Geringer, J.M. (1991) Temporal Discrimination of Modulated Intensity in Music Excerpts and Tones. *Journal of Research in Music Education*, **39**, 29-38

Geringer, J.M , & Sogin, D.W. (1988) An Analysis of Musicians' intonation adjustments within the duration of selected tones. *Contributions to Music Education* , **15**, 1-6.

Geringer, J.M , & Witt, A.C. (1985) An investigation of tuning performance and perception of String Instrumentalists. *Bulletin of the council for Research in Music Education*, **81**, 90-101

Heinich, R. (1984) The proper study of Instructional Technology. *Educational Communication and Technology Journal*. **33**, 67-88

Hillstrom, A.P. & Yantis, S. (1994) Visual Motion and attention capture. *Perception and Psychophysics*. **55**, 435-444

Huron, D. (1991) The Ramp Archetype: A score-based study of musical dynamics in 14 piano composers. *Psychology of Music*, **19**, 33-45

Huron. D. (1992) The Ramp Archetype and the Maintenance of Passive Auditory Attention. *Music Perception*. **10**, 83-92

Jonides, J. & Yantis, S. Uniqueness of Abrupt visual onset in capturing Attention. *Perception & Psychophysics*. **43**, 346-354

Kendall, R.A., & Carterette, E.C. (1990) The communication of Musical Expression. *Music Perception*, **8**, 129-164

Kerr, S. (1989) Technology, Teachers, and the Search for Social Reform. *Educational Technology, Research and Development*. **37**, 5-17

Kuhn, T.L. (1975) Effects of notational values, age, and example length on tempo performance accuracy. *Journal of Research in Music Education*. **23**, 203-210

Kuhn, T.L., & Gates E. (1977) Effects of dynamics, halves of exercise and trial sequences on tempo accuracy. *Journal of Research in Music Education*. **25**, 222-227

Kuhn, P. & Williams, K. (1991) Cross-Modal Speech in Adults and Infants Using Nonspeech Auditory Stimuli. *Journal of Experimental Psychology; Human Performance and Perception*. **17**, 829-840.

Krumhansl, C.L. & Iverson, P. (1992) Perceptual Interactions between Musical Pitch and Timbre. *Journal of Experimental Psychology; Human Performance and Perception*. **18**, 739-751

Liberman, A. & Mattingly, I. (1985) The Motor Theory of Speech Perception revised. *Cognition*. **21**, 1-36.

Lepper, M.R.(1985). Microcomputers in Education: Motivation and social issues. *American Psychologist*, **40**, 1-18.

Madsen, C.K. & Geringer, J.M (1981) Discrimination between Tone Quality and Intonation in Unaccompanied Flute/Oboe Duets. *Journal of Research in Music Education*. **29**, 305-313

Massaro, D.W. & Cohen, M.M. (1983) Evaluation and Integration of Visual and Auditory Information in Speech Perception. *Journal of Experimental Psychology: Human Perception and Performance*. **9**, 753-771

McGrath, M., & Summerfield, A.Q. (1985) Intermodal timing relations and audio-visual speech recognition by normal hearing adults. *Journal of the Acoustical Society of America*. **77**, 533-543

Macdonald, J., & McGurk, H. (1978) Visual Influences on Speech Perceptual Processes. *Perception & Psychophysics*. **24**, 253-257

Macdonald, J., & McGurk, H. (1976) Hearing lips and seeing Voices. *Nature*. **264**, 746-748.

Mathews, M.V. (1969) *The technology of Computer Music*.
Cambridge, The M.I.T.Press

Nakamura, T. (1987) The communication of dynamics between musicians and listeners through musical performance. *Perception & Psychophysics*. **41**, 525-533

Negus, K. (1992) *Producing Pop: Culture and Conflict in the Popular Music Industry*. N.Y. Routledge, Chapman and Hall, Inc.

O'Leary, A., & Rhodes, G. (1994) Cross-modal effects on visual and auditory object perception. *Perception and Psychophysics*. **35**, 565-569

Petkovitch, M.D. & Tennyson, R.D. (1984) Clark's "Learning from Media": A critique. *Educational Communication and Technology Journal*, **32**, 233-241.

Rentz, E. (1992) Musicians' and Non musicians' Perception of Orchestral Instrument Families. *Journal of Research in Music Education*, **40**,185-192

Ross. M., & Morrison.G. (1991) In Search of a Happy Medium in Instructional Research: Issues Concerning External Validity. Media Replications and Learner Control. *Educational Technology, Research and Development*. **37**,19-33

Rosenblum, L.E. & Fowler, C.A. (1991) Audiovisual Investigation of the Loudness-Effort Effect for Speech and Nonspeech Events. *Journal of Experimental Psychology: Human Perception and Performance*. **17**, 976-985

Saldana, H.M. & Rosenblum, L.A. (1993) Visual Influences on auditory pluck and bow judgements. *Perception and Psychophysics*. **54**, 406-416

Schafer, R.M., (1977) *The Tuning of the World*. Alfred A Knopf, New York.

Schonberg, H. *The Lives of the Great Composers*. Futura Publications, London

Solomon, G. Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and Attributions. *Journal of Educational Technology*, 74, 647-658

Ullman, E. (1987) High-Tech Instructional Development: It's the Thought that Counts. *Educational Technology, Research and Development*. 37, 94-106

Walker, A.R. (1987) The effects of musical training on choices of visual metaphors for sound. *Perception & Psychophysics*. 42, 491-502

Walker, A.R. (1990) *Musical Beliefs, Psychoacoustic, Mythical, and Educational Perspectives*. Teachers College Press, N.Y.

Webb, G. (1991) Epistemology, Learning, and Educational Technology. *Educational and Training Technology International*. 28, 120 -128

Appendices

CONCERTO SAINT MARC: Experiment #1 Notation (original)

Grave

T. Albinoni

1

2

3

4

5

6

7

8

9

10

11

12

CONCERTO SAINT MARC: Experiment #2 Notation

Grave

1 2

edit #1 edit #2

Musical staff 1: Treble clef, key signature of two flats, 4/4 time. Measures 1-2. Measure 1 contains a series of eighth notes with a slur. Measure 2 contains a series of eighth notes with a slur and a fermata over the final note. There are 'x' marks above the notes in measures 1 and 2.

3 4

edit #3

Musical staff 2: Treble clef, key signature of two flats, 4/4 time. Measures 3-4. Measure 3 contains a series of eighth notes with a slur. Measure 4 contains a series of eighth notes with a slur and a fermata over the final note. There are 'x' marks above the notes in measures 3 and 4.

4 5

edit #4 edit #5 edit #6

Musical staff 3: Treble clef, key signature of two flats, 4/4 time. Measures 5-6. Measure 5 contains a series of eighth notes with a slur. Measure 6 contains a series of eighth notes with a slur and a fermata over the final note. There are 'x' marks above the notes in measures 5 and 6.

6 7

edit #7 edit #8

Musical staff 4: Treble clef, key signature of two flats, 4/4 time. Measures 7-8. Measure 7 contains a series of eighth notes with a slur. Measure 8 contains a series of eighth notes with a slur and a fermata over the final note. There are 'x' marks above the notes in measures 7 and 8.

8 9

edit #9 edit #10

Musical staff 5: Treble clef, key signature of two flats, 4/4 time. Measures 9-10. Measure 9 contains a series of eighth notes with a slur. Measure 10 contains a series of eighth notes with a slur and a fermata over the final note. There are 'x' marks above the notes in measures 9 and 10.

10 11

edit #11 edit #12

Musical staff 6: Treble clef, key signature of two flats, 4/4 time. Measures 11-12. Measure 11 contains a series of eighth notes with a slur. Measure 12 contains a series of eighth notes with a slur and a fermata over the final note. There are 'x' marks above the notes in measures 11 and 12.

12 13

edit #13

Musical staff 7: Treble clef, key signature of two flats, 4/4 time. Measures 13-14. Measure 13 contains a series of eighth notes with a slur. Measure 14 contains a series of eighth notes with a slur and a fermata over the final note. There are 'x' marks above the notes in measures 13 and 14.

CONCERTO SAINT MARC: Experiment #3 Notation

Grave

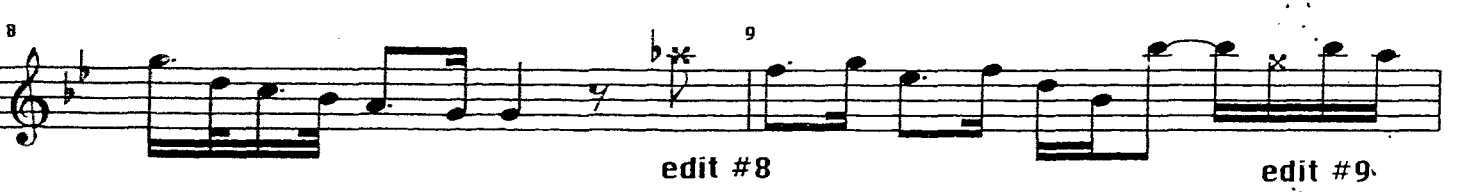
T. Albinoni

1 

3 

4 

6 

8 

10 

12 