

# Effect of Degraded Pitch Cues on Melody Recognition

by

Jung-Kyong Kim

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment  
of the requirements of the degree of Masters of Science

Psychology Department  
McGill University  
Montréal, Quebec, Canada  
June, 2003

mcl

2022757

Copyright © Jung-Kyong Kim, 2003

## Abstract

Past studies of object recognition in vision and language have shown that (1) identification of the larger structure of an object is possible even if its component units are ambiguous or missing, and (2) contexts often influence the perception of the component units. The present study asked whether a similar case could be found in audition, investigating (1) whether melody recognition would be possible with uncertain pitch cues, and (2) whether adding contextual information would enhance pitch perception. Sixteen musically trained listeners attempted to identify, on a piano keyboard, pitches of tones in three different context conditions: (1) single tones, (2) pairs of tones, and (3) familiar melodies. The pitch cues were weakened using bandpass filtered noises of varying bandwidths. With increasing bandwidth, listeners were less able to identify the pitches of the tones. However, they were able to name the melodies despite their inability to identify the individual notes. There was no effect of context; whether or not listeners heard single tones, pairs of tones, or melodies did not influence their pitch identification of the tones. Several possible explanations were discussed regarding types of information that listeners had access to, since they could not have relied on detailed features of the melodies.

## Résumé

Les études antérieures portant sur la reconnaissance des objets visuels et linguistiques ont montré que (1) l'identification d'une structure plus large d'un objet est possible même si ses unités composantes sont ambiguës ou manquantes, et (2) les contextes influencent souvent la perception des unités composantes. La présente étude cherche à déterminer si un cas similaire peut être trouvé pour les objets auditifs, examinant (1) si la reconnaissance de la mélodie peut être possible avec des indices de hauteurs incertains, et (2) si l'ajout d'information contextuelle peut améliorer la perception de la hauteur. Seize auditeurs possédant une formation musicale ont tenté d'identifier, sur le clavier d'un piano, les hauteurs de sons dans trois conditions de contexte différentes : (1) sons isolés, (2) paires de sons, et (3) mélodies familières. Les indices de hauteur étaient affaiblis via l'utilisation de bruits blancs filtrés par des filtres passe-bande de différentes largeurs de bande. Avec l'accroissement de la largeur de bande, les auditeurs étaient moins capables d'identifier les hauteurs des notes. Néanmoins, ils étaient capables de nommer les mélodies malgré leur incapacité à identifier les notes individuelles. Le contexte n'a pas eu d'effet ; que les participants aient entendu des sons isolés, des paires de sons, ou des mélodies, n'a pas influencé leur identification de la hauteur des sons. Plusieurs explications possibles sont proposées en rapport avec les types d'information auxquels les auditeurs avaient accès, étant donné qu'ils n'auraient pas pu se baser sur les caractéristiques détaillées des mélodies.

## Acknowledgements

I would like to acknowledge several people who have contributed to shaping of the present research paper. I would like to acknowledge the supervision of Dr. Daniel J. Levitin during the design, execution, and preliminary write-up of this research. I thank him for many hours of discussions throughout the course of this study, and I am also grateful for his advice on writing. I would also like to acknowledge Dr. Albert S. Bregman in the final write-up. I appreciate his comments on theoretical implications of the experimental results. He helped me to strengthen the ideas that I had from the study and motivated me to think about my research in a larger context. I also thank him for giving me immeasurable advice and support. I would also like to acknowledge Dr. Roger Sheppard who contributed to the conception of the present study. I wish to thank Caroline Traube who designed the computer program for the stimuli, and who helped me immensely with the creation of the stimuli and translation of the abstract to French. I would also like to thank Dr. Regina Nuzzo for her statistical advice, and most of all her mental support and friendship. I would like to thank Rhonda Amsel and Dr. Mark Baldwin for giving me valuable advice. I am grateful to the Department of Psychology of McGill University for its support for during the course of the past two years. Finally, but not least, I would like to thank my sisters and my parents for their immeasurable love and emotional support.

The data used in the present study were presented at the meeting of the Canadian Acoustical Association in October of 2003 and were published in a proceedings paper [Kim, J., & Levitin, D. J. (2002). Configural processing in melody recognition. *Canadian Acoustics*, 30, 156-157]. This research was supported in part by a grant from the National Sciences and Engineering Council of Canada to Dr. Daniel J. Levitin.

## Table of Contents

Abstract .....	ii
Résumé.....	iii
Acknowledgements.....	iv
List of Figures & Tables .....	vi
INTRODUCTION .....	1
Environment for object recognition .....	2
Superiority of recognizing a larger unit over its component units.....	3
Object perception .....	4
Language .....	11
The present study .....	20
Review of melody perception studies .....	21
Rationale for the present study.....	25
METHOD.....	28
Participants.....	28
Materials & Apparatus .....	28
RESULTS .....	33
Percentage of correct identification .....	33
Semitone errors .....	40
DISCUSSION .....	47
Access to contour information .....	48
Access to interval information .....	50
Approximation of interval sizes.....	51
Interaction between different elements .....	51
Did context improve pitch perception?.....	53
REFERENCES.....	57
Appendix A .....	62
Appendix B .....	63

## List of Figures &amp; Tables

Figure 1. Percentage of accurate pitch identification.....	35
Figure 2. Percentage of accurate melody & pitch identifications in Melody condition. ....	37
Figure 3. Percentage of accurate pitch identification in Double-tone & Interval conditions. .....	39
Figure 4a. Frequency distributions of errors of different magnitudes for Single-tone condition.....	41
Figure 4b. Frequency distributions of errors of different magnitudes for Double-tone condition.....	42
Figure 4c. Frequency distributions of errors of different magnitudes for Interval condition. .....	43
Figure 4d. Frequency distributions of errors of different magnitudes for Melody condition. .....	44
Figure 5. Mean size of semitone errors as a function of bandwidth. ....	46
Table 1. Bandwidths applied to filtered white noise.....	30
Table 2. Percentage correct for pitch identification.....	34
Table 3. Mean size of semitone errors for pitch identification .....	45

## INTRODUCTION

The present thesis concerns object recognition in audition. The main issue deals with the question of how listeners recognize familiar melodies when the individual elements of the melodies are ambiguous. I ask the question of whether holistic processing exists in melody recognition. I chose a melody as an auditory object to be recognized because a familiar melody can be a good candidate for a Gestalt object in audition. Dewitt and Crowder (1986) mentioned, “melodies could theoretically be heard as just a series of individual pitches, but there are some more global features of melodies that seem to allow (or compel) us to group these individual pitches together and recognize the melody as a unit” (p.259). If melodies are considered as perceptual units (i.e., melodies are more than individual pitches organized together, but have emergent properties as a result of arranging individual pitches in a certain way), it is possible that melodies may be recognized even if identities of the individual pitches are made ambiguous. Even though component units are uncertain, recognition of the larger structure may still be accomplished through holistic processing.

The introduction divides into three main sections. The first section will discuss the variance of environments in which objects are recognized, and the perceptual system’s flexibility in compensating for the interference of the environment by mean of global processing as opposed to local processing. The second section will provide a review of studies that reported superiority of recognizing a larger unit over its component units in object, face, and language perceptions. The review of object recognition will discuss the object superiority effect, configural superiority and orientation effects, and visual agnosia, all of which demonstrate that objects can be recognized even if individual

features of the objects are ambiguous. The review of face perception will include topics such as configural processing and prosopagnosia which show that global structure of a face can be processed separately from the analysis of local features. The review of language perception will examine phenomena such as the word superiority effect, the phonemic restoration effect, phonemic transformation of steady-state vowels, and sine wave speech. They demonstrate that units of language might be processed in an integral fashion as opposed to independent processing of individual units.

### **Environment for object recognition**

It is worthwhile to ask at this point why it would be necessary that our perceptual system might acquire a mechanism that causes whole perception to be greater than the sum of part perceptions. Our environment is not always optimal for object recognition since we encounter so much interference. For example, it is very rare that a listener hears a speech without any interfering background noises. It is rare that we see an object with all of its form intact. Parts of the objects are often occluded by other objects. The same condition applies to face perception. It is rare that we see a face in the same condition all the time. Sometimes, we only see parts of a face (because hair is covering up one part of the face, or another object is occluding a part of the face). Also, objects are rarely seen from the same angle or heard from the same distance. Our environment creates so much interference that it is rare that we encounter objects in a single way. However, we are able to recognize them as invariant objects. When we talk to a person at a distance or over the phone, we know that the qualities of the voices change, but we still perceive them to be the same person's voice. We still recognize a person's face as the same person's face even if we see it from many different angles. Our sensory system is



adaptive to changes and able to sort out what is part of an object and what is not. Since we hardly encounter objects in their perfect physical conditions, it perhaps is not necessary that we process detailed features of the objects. Maybe it is important that we are able to perceive objects at a global level in the sense that object recognition is still possible without analyzing all the low-level features (i.e., bottom-up processing), and that we have a mechanism that can compensate for the flexibility of the environment. From this argument, we see that holistic or global processing of an object is just as necessary as feature-based or local processing. I would like to provide an overall literature review of cases of object recognition in which holistic processing dominates over processing based on local features. I will present examples of the superiority of recognizing a larger unit over its components in object and face recognition in visual processing, and in reading and speech in linguistic processing. Furthermore, I will address how these examples relate to melody recognition, the main topic of my investigation.

### **Superiority of recognizing a larger unit over its component units**

A large number of studies in the areas of visual perception have reported that even if component units of an object are uncertain, the larger unit can still be recognized. Examples can be found in object perception in which visual elements are sometimes detected accurately or easily when they are placed in certain contexts, and in face perception in which faces can be recognized most easily when facial features are placed in the appropriate facial context.

## **Object perception**

### *Object superiority effect*

The object superiority effect (Weisstein & Harris, 1974) is an example in which people are able to identify a component unit much more accurately if it is embedded in the form of an object rather than in isolation. Subjects in Weisstein and Harris's experiment were shown a target line followed by four diagonal lines differing in orientation and spatial location. The lines following the target line could be presented in various contexts. One of the contexts was that the lines were embedded in three-dimensional objects. Other contexts were such that the lines were part of a collection of lines arranged in different configurations that did not yield three-dimensional images. The task was to choose the stimulus that contained the target line as fast as possible. Subjects were more accurate in finding the correct diagonal lines when they were embedded in three-dimensional objects rather than when they were part of other line configurations. Weisstein and Harris suggested that it is the well-structured pattern of an object that enabled subjects to detect the embedded line, since the patterns that did not form a unitary object did not yield fast detection of the line. Weisstein and Harris further suggested, "perhaps recognition of ... objects depends on more general processes that make use of structural rules and meaning to determine perception" (p.754). A constituent element (line segment) is perceived better when the context creates a well-formed unit, even if the context provides no clues about the correct choice on a given trial, and even though the identity of the whole configuration depends on the identity of its components (line segments).

### *Configural superiority effect*

Configural superiority effect is another example of the superiority of recognizing a larger unit over recognizing one of its components. Pomerantz, Sager, and Stoevers (1977) showed subjects four stimuli for a brief moment. Among the four, three were the same and the other one was the odd stimulus. In one condition, the stimuli were diagonal lines, with one oriented in one direction and the other three oriented in another direction. Subjects were asked to find the odd line that pointed in a different direction from the three other lines. In another condition, the diagonal lines were embedded in a configuration. For one stimulus, a diagonal line was a hypotenuse of the right angle triangle, and for the other stimulus, (instead of being a hypotenuse) a diagonal line projected from two lines that were at the right angle from each other, therefore forming an arrow. Subjects then were asked to find the stimulus that had a different configuration from those of three other stimuli. Pomerantz et al. found that the odd line was found much faster when the lines were part of either the arrow or the right triangle rather than when they were presented in isolation. This finding suggests that lines are more discriminable if they are placed in a certain context than if they are presented without any context. In other words, it was easier to detect differences among compound structures rather than differences among its simpler components in isolation. The configurational advantage also suggests that in order to recognize the lines that were placed in the context faster than the single lines, subjects might have processed the configuration as a single, global unit rather than processing local elements of the stimuli separately. Pomerantz et al. argued that a whole is different from the sum of its parts because parts, when they are placed in a certain context, interact with one another to yield emergent features (e.g., intersections of the

arrow or closedness of the right triangle). Furthermore, our perceptual system treats the emergent properties as functional perceptual units.

### *Configural orientation effect*

Similar to the configural superiority effect, the superiority of recognizing a larger unit over recognizing its component units can also be found in the perception of spatial orientation of equilateral triangles. Palmer (1980, 1999) demonstrated that the perceived directional orientation of equilateral triangles depended on their configural arrangements. An equilateral triangle has an ambiguous pointing orientation when it is presented by itself. However, when several equilateral triangles are aligned, they are perceived to point in a direction. For example, when the triangles are aligned by their axes, they are perceived to point toward the direction of the axis. When the triangles are aligned by their bases, they are perceived to be oriented perpendicular to the base. Palmer argued that the overall configuration of the figures could strongly influence the direction of the ambiguous figures such as equilateral triangles. This is an example of how a global pattern influences the perception of its component units even though the component units have uncertain properties by themselves. (It may, in fact, require that the component patterns have some ambiguity of interpretation before the biasing can work.)

### *Visual agnosia*

Consistent with the evidence that individual features of a visual object (whether it be a face or a geometrical shape) can be ambiguous but the object can still be recognized, or can be identified faster than its parts, there are examples of visual impairments in object or face recognition that support the notion of processing of a global unit that is separate from processing of its component units. “Visual agnosia” (Humphreys &

Riddoch, 2001) is a visual impairment that prevents people from recognizing visual objects. Humphreys and Riddoch describe a particular agnosic patient, HJA, who suffered a stroke that impaired his recognition of objects. HJA's impairment was not due to a difficulty in encoding basic shape features nor was his problem due to a loss of stored knowledge of objects. He was able to copy drawings of objects, implying that his ability to analyze basic features was reasonably intact. Had he not have an understanding of the basic features of the objects, he would not have been able to copy the objects. Also, he was able to provide good verbal definitions for objects, suggesting that his long-term memory for objects were intact. The problem lay in his difficulty in integrating basic features to recognize an object. For example, he would describe a paintbrush as "two objects lying close to one another" (p.209), but was unable to integrate the basic features of the object to perceive as a global shape, a paintbrush. The agnosic patient was able to "see" but unable to "recognize" objects because he could not process the objects as wholes despite his ability to process their component units. Even though he was able to sum parts of the object, he could not recognize the summed product (the whole). This disorder could support the Gestalt notion of "whole is greater than sum of its parts" because agnosic patients are able to process basic features of an object but lack the ability to integrate the features into a whole.

### *Global/local information processing*

In addition to the symptoms of agnosic patients, there is evidence that global and local information in vision is processed independently. Delis, Robertson, and Efron (1986) found that people with right hemisphere damage show deficits in global processing but intact local processing, and people with left hemisphere damage show the

opposite deficits. When asked to copy a structure such as a letter or a triangle consisting of smaller units such as letters or rectangles, the right hemisphere damaged patients were unable to draw the whole structure although they could draw the component units. On the other hand, the left hemisphere damaged patients were able to follow the whole structure although they could not reproduce the component units. The separate hemispheric deficits in perceiving the global and local information imply that the two types of information might be processed independently. If a global structure can be perceived independently of its component structure, then this implies that it is possible that objects can be perceived at a global level without attending to the component elements of the global structure. It should be noted that this is a very special type of stimulus. In this type, the details of the local structure do not contribute to the global structure and vice-versa; they are independent. However, in melodies, the components (notes) do contribute to the global structure since changing one note changes the melody.

## **Face perception**

### *Configural processing*

Face perception is a well-known example of holistic processing. That is, it is possible that people recognize and discriminate faces without attending to individual elements. Past literature suggests that people do not need to identify individual features of a face in order to recognize the whole unit as a face, or even if the individual features are uncertain, people can still recognize a face as long as the individual features are placed in the appropriate context of a face. Palmer's (1975) fruit face and a line face are simple demonstrations. Palmer showed that a certain configuration of fruits could allow recognition of a face despite the literal identities of the local features. In fact, the Italian artist from four centuries earlier, Guiseppi Arcimboldo (1527-1593), had already

demonstrated the same principle by painting faces and busts whose features consisted of fruits, vegetables, flowers, fish, and other objects (Effetto Arcimboldo, 1987). Fruits, if presented in isolation, would not be considered as individual features of a face. However, a certain arrangement of the fruits in Palmer's research (e.g., a watermelon in the global face position, two apples in the eye positions, a pear in the nose position, and banana in the mouth position) allowed recognition of a face despite the literal identities of the local features. Also, Palmer showed that simple lines are sufficient to be recognized as faces, as long as the lines representing individual features of the face are arranged in appropriate places within the context of the face. If the individual features of the face were represented in isolation from a facial context, a greater detail would be required for recognition of the features. Consistent with Palmer's argument of configural processing in face recognition, it has been reported that a face can be detected much more easily when it is in the upright position rather than when it is inverted, or the face features rearranged into a meaningless configuration (Purcell & Stewart, 1988; Homa, Haver, & Schwartz, 1976).

Recent findings also suggest that people use configural information for identifying faces rather than processing parts or features to build up a face representation (McKone, Martini, & Nakayama, 2001; Moscovitch, Winocur, & Behrmann, 1997; Tanaka & Farah, 1993). McKone et al. defined "feature-based" face identification as using local arrangement of facial elements in order to identify a face. They eliminated the possibility of feature-based identification by isolating the configural processing by means of degrading the local features of their stimuli with white noise. When individual features of the faces were obscured, it was the configuration of the particular faces that people relied on in identifying them. When the configural pattern was violated (e.g., only a single

feature, nose of a face was presented, or a face was inverted), people's identification (as measured by categorical perception) ability was very low. This suggests that people did not necessarily use detailed, local features but that they instead relied on the overall configuration of a face in making identifications.

### *Prosopagnosia*

Similar to the object recognition disability in agnosic patients, an equivalent impairment is also known to exist in face recognition. "Prosopagnosia" is a disorder of face recognition (Humphreys & Riddoch, 2001, p.213). The symptoms of prosopagnosic patients are similar to those of agnosic patients in that they can accurately describe individual features of a person's face yet are unable to recognize the face even if it is a face of someone that they know well (Palmer, 1999; Humphreys & Riddoch, 2001). Therefore, it is possible that face perception involves processing of a global representation of a face and the holistic process might occur in parallel to the feature-based analysis (also see Saumier, Arguin, & Lassonde, 2001).

Taken together, a large collection of past literature in object and face perception supports the Gestalt notion of "a whole is greater than (or different from) sum of its parts". The literature suggests that it is possible that the global structure of an object or a face can be processed separately from the analysis of local features. Now I would like to turn my attention to the linguistic domain where examples of superiority of a larger unit over its component units can be found. Striking examples of how a larger unit is analyzed with its component units missing or ambiguous can be found in speech perception where context plays a crucial role.



## Language

### *Word superiority effect*

The word superiority effect (Reicher, 1969) is a good example that demonstrates the superiority of the holistic processing of a global structure over the processing of its component units. The subjects in Reicher's study were given a brief tachistoscopic display of one letter or one four-letter word. After the presentation, they were given two optional letters and asked which letter they had seen from the display. If they were presented a word, then they were asked to choose a letter that they had seen in the word. For example, subjects could have seen a letter D, and would have chosen from either D or K. In the word condition, they could have seen a four-letter word such as WORD, and they had to choose whether they had seen a letter D or K in the word. In this case, the incorrect letter K would have had to make up a word (e.g., WORK) if it was embedded in the place of the correct letter. Reicher found that subjects were more accurate in identifying the correct letter if it was presented in a word than as a single letter. This finding implies that given the very short duration of time, subjects must have processed a word faster than a single letter although there are four times as many letters to process in a four-letter word than a single letter alone. In other words, if processing of a four-letter word is equivalent to processing of four single letters, it should take longer to process a word than a single letter. However, this was not the case. Furthermore, since the subjects did not know which letter position they were going to be tested on, they must have recognized all four letters of the word with an equal probability. Therefore, Reicher suggested that word perception involves more than identification of individual letters, and that the perceptual unit of a word is larger than a single letter. Wheeler (1970) also found

the word superiority effect using similar methods and supported Reicher by suggesting, “there is an interaction among the letters such that the context of the other letters of a meaningful word improves recognition” (p.78). Therefore, not only does there seem to be a holistic processing of a larger unit, but the context of the larger unit also seems to help identify the component units. The word superiority effect seems to be applicable to words only, since the effect did not take place with identification of letters of nonwords (Reicher, 1969; Allegretti & Puglisi, 1982). Findings of other studies support the holistic processing of words independent of letter coding (Johnson, 1975; Johnson & Marmurek, 1978; Jacewicz, 1979; Lawry & LaBerge, 1981; Chastain, 1982).

### *Phonemic restoration*

The phonemic restoration effect is an example in the linguistic domain that demonstrates analysis of a larger unit with a component unit missing. It is a phenomenon in which people hear an utterance in which a phoneme is deleted and replaced by a non-speech sound, and the linguistic contexts aids or biases identification of the missing speech sound. Warren (1970) discovered that when a phoneme in a speech such as “the state governors met with their respective legislators convening in the capital city” (e.g., the /s/ in the word “legislators”) was deleted and replaced by a cough, people still heard the missing phoneme and were unable to accurately locate at which place in the sentence they had heard the cough. It was as if people had heard both the cough and the missing phoneme. Even when the cough was replaced by a 1000Hz tone, people reported that they heard the missing phoneme. In the extension of Warren’s study, Warren and Warren (1970) presented sentences in which the first phoneme of a word ending in “eel” was missing. In the following examples the asterisk represents the missing phoneme that was

replaced by a cough: “it was found that the \*eel was on the axle”, “it was found that the \*eel was on the shoe”, “it was found that the \*eel was on the orange”, or “it was found that the \*eel was on the table.” Depending on the sentences heard, people reported hearing “wheel”, “heel”, “peel”, or “meal” respectively. Warren (1970) suggested that we often listen to speech against background noises, and phonemic restoration is a mechanism that our auditory system has developed in order to compensate for any sounds that are lost due to the extraneous noises in our environment. The results from the two experiments suggest that our auditory system is able to “restore” a phoneme that does not physically exist at all as long as there is enough evidence to *infer* that the phoneme could have been masked by the non-speech sound. Linguistic context (e.g., the meaning of the sentence) plays an important role in “disambiguating” the missing sound because it is the context of the sentence that provides a clue about the sound that might be present. It is almost as if the auditory system is looking for the particular phoneme in the non-speech signal based on what it knows about the sentence, and restores it if there is enough information to make such an inference. In Warren’s study (1970), if the phoneme was not replaced by any sound (i.e., there was a silent gap in replacement of the phoneme), people were able to notice the gap in the speech. The fact that the silent gap did not yield phonemic restoration suggests that the auditory system is equipped to interpret only if there is a signal that can lead to infer a missing sound. It is also interesting to note that just as in vision, the auditory system employs top-down processing in order to fill in missing information in the auditory scene, and it is possible to recognize the larger structure of speech even though there is uncertainty in the component units.

### *Phonemic transformation of steady-state vowels*

Another striking example of our auditory system's ability to interpret ambiguous component units using top-down processing is the "phonemic transformation effect" of steady-state vowels. The basic idea behind the phenomenon is that when steady-state vowels are repeated in a loop, listeners tend to hear words or syllables. The phenomenon was first discovered when Warren, Bashford, and Gardner (1990) presented subjects a repeated sequence of three vowel sounds (e.g., /ʌ/ as in "hud", /æ/ as in "had", and /i/ as in "heed") that were recorded at the same fundamental frequency. When the sequence was played at a speed between 30 and 100msec per vowel, instead of hearing steady state vowels (as they were), subjects reported hearing words or syllables that did not resemble the actual vowel sounds at all. According to Warren et al., this was an interesting finding because normal conversation has an average duration of speech sounds of approximately 80-100msec, and this duration was similar to the rate found to generate the phonemic transformation effect. Warren et al. mentioned that people tend to have a much higher threshold than 100msec for identifying the order of component items in the recycled sequence of sounds (i.e., longer duration per sound item is required for identification of order). Nonetheless, speech sounds occur at a rate faster than that at which identification of order is possible, and at this rate steady-state vowels were organized into words and syllables. Warren et al. suggested that the auditory system uses its lexical knowledge to interpret the steady-state vowel sequences by matching the auditory signal to the lexical template that we have developed. The lexical template would be the knowledge for word composition rules (i.e., combinations of sounds that can or cannot make up words) in the language that the subjects are familiar with. Warren et al. argued that we do not need to

know the order of phonemic sounds because the way we identify speech is not through identification of its constituent phonemes. Instead, the auditory system treats the patterns that are formed by different arrangements of acoustic signals as “temporal compounds”. Considering these temporal compounds to be perceptual units, our auditory system does not require identification of component speech sounds in order to recognize words or discourse. This relates to my discussion of holistic processing in recognition because speech perception does not seem to require analytic processing of individual phonemes. Instead, it seems to employ a holistic analysis of the sounds of words. Further experiments with sequences of four steady-state vowels confirmed the holistic processing of speech sounds (Chalikia & Warren, 1991; Chalikia, & Warren, 1994; Warren, Healy, & Chalikia, 1996). This is an example of the superiority of recognizing a larger unit over its component units.

#### *Lexical influence on phonemic processing*

Consistent with the phonemic transformation effect found by Warren et al. (1990), evidence from studies of phonetic processing (Tomiak, Mullennix, & Sawusch, 1987; Ganong, 1980) also seem to suggest that speech perception is more than a mere sum of acoustical signals. The study by Tomiak et al. (1987) demonstrated that we perceive phonemes within a syllable in an “integral” manner rather than treating each phoneme as an independent unit. Subjects in Tomiak et al.’s study were instructed that they would hear a sequence of a noise and a tone. They were told in advance that the target was either a noise or a tone, and upon hearing the sequence, they were to classify the target as quickly as possible. On the other hand, another group of subjects in the same study were told that the stimuli were syllables each consisting of a fricative and a vowel, and their

task was the same as for the other group (to classify the target phoneme as quickly as possible). The stimuli that were used were syllables such as /fæ/, /jæ/, /fu/, and /ju/, and other than instructions given, exactly the same stimuli were used for the two groups of subjects. Results indicated that the reaction time of the subjects who treated the acoustical sequence as speech was substantially greater than that of the subjects who treated the sequence as non-speech. The shorter reaction time for the “non-speech” subject group indicates that the noise and tone could be processed independently. On the other hand, the longer reaction time for the “speech” subject group suggests that phonemes within a syllable had to be initially processed in an integral fashion, and thus, it would take longer to decompose the syllable into individual phonemes. The difference in the reaction time demonstrates that speech processing employs a mode that treats a syllable as a holistic unit rather than the mere addition of two phonemes (also see Day & Wood, 1972; Wood & Day, 1975).

Ganong (1980) used a different approach to arrive at a similar conclusion to that of Tomiak et al. (1987) that syllabic perception is more than serial processing of the sequence of phonemes. Ganong extended the studies of categorical perception in speech (Lisker & Abramson, 1970; Eimas & Corbit, 1973). Categorical perception of speech generally refers to the phenomenon in which there is a sharp boundary in perception when varying the properties of a speech sound gradually between the values that define two unambiguous speech sounds. An example is the boundary between voiced and unvoiced phonemes (e.g., /b/ and /p/) when the voice onset time (VOT) that characterizes one phoneme is gradually increased or decreased to yield the other phoneme. Despite the gradual change in VOT, people keep hearing one phoneme until they start to suddenly hear the other phoneme (i.e., there is no perception of intermediate phonetic sounds). The

VOT at which perception changes from one to the other phoneme is called the phonetic boundary. Ganong prepared a pair of monosyllabic words, one containing a voiced phoneme and the other containing an unvoiced phoneme. Word selection was such that one of the pair was a real word (e.g., beef), and the other was a non-word (e.g., peef). Then, the VOT of the phonemes were manipulated in order to produce different syllables with varying VOTs. Consistent with other studies, Gagnong replicated the categorical perception (i.e., people either heard “beef” or “peef”). However, the phonetic boundary (the VOT at which perception of the word changes from one word to the other) was shifted toward the non-word VOT in comparison to the phonetic boundary when two neutral words were at either end. This result implies that people tended to hear the real word more than the non-word in the VOT continuum, demonstrating that lexical knowledge biases our phonetic categorization. In other words, certain VOTs that would have resulted in perception of /p/ on the /b/-/p/ continuum would result in perception of /b/ on the /beef/-/peef/ continuum. If each phoneme of the real word is treated in the same way as each phoneme of the non-word, then the phonetic boundary should not have been shifted. However, the change in phonetic boundary suggests that phonemes in real words are processed in a different way than those in non-words, or at least that some correction is applied once the word identity becomes available. As in the phonemic restoration effect, there seems to be top-down processing (i.e., influence of lexical knowledge) that affects the perception of individual segments of speech. The shift in the phonetic boundary further suggests that phonemes of words could be treated as an integral unit because if each phoneme is processed as an independent unit, then perception of the phoneme should not have been affected either. The examples of integral processing of syllables demonstrate that perception of syllables might not

necessarily require prior analysis of their phonemes. It is possible that the larger structure is acoustically analyzed independently of or in parallel to the analysis of its component units.

### *Mathematical expressions of integral processing*

Having shown that perceptual units for speech are larger than single phonemes, two studies developed mathematical formulas that express the relationship between perception of a larger unit and perception of its component units. Boothroyd and Nittrouer (1988) measured the effects of context on the perception of speech in noise, and was able to obtain the equation,  $p_w = p_p^j$ , where  $p_w$  is the probability of recognition of wholes,  $p_p$  is the probability of recognition of constituent parts, and  $j$  is the effective number of statistically independent parts within a whole. Boothroyd and Nittrouer found that for nonsense consonant-vowel-consonant (CVC) words,  $p_w = p_p^3$ , meaning nonsense words are perceived as though they consist of three independent units. However, for CVC words,  $p_w = p_p^{2.5}$ , indicating that three-letter words are perceived as though they consist of 2.5 independent units. The  $j$  value of 2.5 (instead of 3) indicates that identification of each phoneme is superior when phonemes are grouped in a meaningful word than a non-word. Also, Boothroyd and Nittrouer found the  $j$  value to range between 2.5 to 1.6 for recognition of four-word sentences that are highly predictable. This indicates that instead of treating four words as four independent units, people tend to perceive the four-word sentences as though they consist of 1.6 to 2.5 independent units. Another study (Versfeld, Daalder, Festen, and Houtgast, 2000) reported similar values for sentence recognition for four-word sentences; the  $j$  value ranged from 1.5 to 2.54. Versfeld et al. suggested that because words in a meaningful sentence are “redundant” in



their meanings, the effective number of words are fewer than four ( $j = 4$  indicates that four words are treated as though they are independent units). Versfeld et al. further suggested that it is redundancy in speech that allows flexible judgment about the identity of the speech despite extraneous sounds that interfere with the speech.

### *Sine wave speech*

Perhaps the most striking example of integral processing in speech perception would be sinewave speech. A sinewave speech signal is a replication of a natural utterance that retains only the coarse-grain changes in the speech spectrum over time (Remez, Rubin, Pisoni, & Carrell, 1981; Remez & Rubin, 1990; Remez & Rubin, 1984). It consists of three or four sinusoids that track the variations of frequency and amplitude of vocal resonances in the natural utterance. However the signal cannot represent the bandwidths of the formants, the glottal pulsations, or any short-term acoustic cues such as transient sounds attributed to consonants. Despite the unnatural characteristics of the sinewave pattern, listeners have reported hearing speech-like sounds or sometimes have even identified the natural utterances that the syntheses were modeled upon. This finding suggests that speech perception might be independent of the component units that compose the utterance. Remez and Rubin (1990) argued that listeners could understand speech simply by extracting a gross pattern of frequency and energy changes from an utterance. The sinewave speech is an example of the notion that perception of a larger unit does not necessarily depend on the exact details of the perception of component units. In this case, perception of component units was not even possible because the component units did not even exist.

## *Alexia & Dyslexia*

Equivalent to visual agnosia as evidence for holistic processing in object recognition, there are disorders in the linguistic domain that support the idea of holistic processing in word recognition. Alexia (Humphreys & Riddoch, 2001) is a disorder that is characterized by difficulty in recognizing words despite intact identification of the component letters. Alexic patients are either unable to recognize words or often take longer than normal readers to recognize words, because they have to read the words letter by letter. They lack the normal reader's ability to treat words on a holistic level.

Dyslexia (Humphreys & Riddoch, 2001) is a disorder that is characterized by symptoms opposite to alexic symptoms. Dyslexic patients have do not have difficulty recognizing words but have problems in identifying individual letters of the words. This impairment supports the notion that word recognition may be possible without accurate identification of individual letters, thus further extending to the possibility that recognition of a larger unit does not necessarily require a prior detailed analysis of its component units.

### **The present study**

Thus far, I have provided numerous examples in the visual and linguistic domains that recognition of a larger unit can be superior to recognition of its component units. The present study investigates whether the same case can be found in audition. The notion of melody recognition goes back to the Gestalt psychologists' original question of object recognition. The Gestalt movement was launched by von Ehrenfels (1890) with the following object identification problem: "How is it that when we hear a melody, for example, one consisting of six tones, and then the same melody transposed to a new key, that we recognize it as the same, even though the sum of the elements is different?"

Ehrenfels answered that there are holistic attributes that emerge from melodies. In other words, a “Gestaltqualität (Gestalt feature), such as a melody, was a feature of the note sequence, over and above the features of the individual tones. Key to the Gestalt approach is that isolating elements from their larger context loses significant, configural properties of the object in the process. Garner (1981, p.119) agrees: “A configuration has properties that have to be expressed as some form of interaction or interrelation between the components, be they features or dimensions.” Furthermore, the Gestalt psychologists believed that the same principle applies to general object recognition. They believed that perception of a whole is greater than the sum of the perceptions of the parts, and that there are emergent features that we perceive when components units form a larger structure. Before explaining details of the present study, let me first review what is known about melody recognition.

### **Review of melody perception studies**

Melodies contain three levels of information. First, melodies can be described by the absolute pitches of their component notes. Second, melodies contain relative pitch (or interval) information. That is, regardless of the absolute pitch information, they contain the precise relative magnitude of the pitches of two adjacent notes. Third, melodies contain contour information (pattern of ups and downs). Contour is inherent in intervals, and intervals are inherent in absolute pitches in the sense that once the absolute pitches of a melody are provided, the intervals and contour pattern can be known. However, it may be that absolute pitches, intervals, and contour can be processed independently. We already know that absolute pitch information and intervals are coded separately because familiar melodies can be transposed into different keys and we can still recognize them.

A number of studies have tried to isolate contour information from intervals, and examine whether contour processing is independent of interval processing in melody recognition.

*White (1960)*

One of the earliest studies on melody recognition was White's study (1960) of melody recognition after various transformations. White manipulated the pitch information of familiar melodies in various ways. Several melody transformations involved distorting the precise relative interval information but preserving relative magnitudes of the intervals were reserved. For example, one of the transformations was to double the size of each interval (e.g., an interval of 4 semitones would be transformed to an interval of 8 semitones). Other transformations involved distorting the precise relative interval information as well as the relative magnitudes of the intervals. For example, in one condition, all the intervals were set to one semitone, but the sign of the original pitch directions was maintained so that only the contour information was preserved. It was found that people were able to recognize, with high accuracy, the melodies whose relative magnitudes, or intervals, were preserved (80% accurate or better). When only the contour information was kept intact, people recognize the melodies with only around 60-69% accuracy. Thus, White demonstrated that melody recognition does not necessarily require exact interval sizes as long as relative magnitudes of the intervals between adjacent notes are preserved. For example, if Interval A was larger than Interval B but smaller than Interval C in the original melody, as long as the interval magnitude was maintained in the decreasing order of Intervals C, A, and B after the transformation, listeners would recognize the distorted melody. Furthermore, melody recognition was

possible using only contour information, indicating that contour might be encoded independently of interval information.

### *Dowling's studies*

Dowling and his colleagues performed a series of studies focusing on melody recognition in terms of contour- and interval-based identifications. They demonstrated that there are different roles that contour and intervals play in melody perception. A series of studies on contour and interval cues in melody recognition suggested that listeners use contours more often than intervals as cues for recognizing novel melodies (Dowling, 1978; also see Dewitt & Crowder, 1986), or for melodies that have an unstable tonal context (Dowling, 1982; also see Dowling, 1991). In contrast, listeners tend to use intervals more often than contours in recognizing familiar melodies (Dowling & Barlett, 1981; Dowling & Fujitani, 1971). Dowling and his colleagues suggested that contour information is immediately available for short-term memory tasks but it might be accessed rather indirectly through interval information after a long delay. Also, memory for intervals is resistant to forgetting once it is stored in long-term memory (also see Dewitt & Crowder, 1986).

Among the studies, one study (Dowling & Fujitani, 1970) examined recognition of familiar folk tunes based on contour information. Listeners were presented distorted versions of familiar folksongs. The melodies could be distorted in such a way that the contour and relative interval size (not the exactly interval size) were preserved, or only the contour information was preserved. It was found that listeners were 66% accurate in recognizing the melodies with their contour and relative interval sizes preserved, and 59% accurate in recognizing the melodies with their contour information preserved. The

chance level was at 20% (since listeners had to choose a correct title from five alternative songs). Consistent with White's (1960) finding, Dowling and Fujitani (1970) found that although exact interval sizes are more important for recognizing familiar songs, it is still possible to recognize distorted melodies based on relative interval information. Also, although relative interval information gives more clues about the melodies (since there was 7% difference in accuracy with or without relative interval information), contour information can be used alone to yield melody recognition.

### *Peretz's studies*

Studies of Peretz and her colleagues proposed that global processing in music is associated with perceiving melodies in terms of contour, and local processing associated with perceiving interval relationships. They suggested that contour-based and interval-based identification mechanisms are separate processes. In the study of a patient with right hemisphere damage (Lassonde et al., 1999), hemispheric dissociations of global versus local processing have been found. In this study, tasks involved discriminating a target melody from an original melody. In one condition, one part of the original melody was manipulated so that the target melody differed from the original melody in intervals but the contour was preserved. In another condition, one part of the original melody was manipulated such that the target melody had a different contour. The patient was able to discriminate melodies that differed in intervals but they could not detect the difference in melodies that differed in contour. Thus, Lassonde et al. suggested that global processing might be dominated by the right hemisphere (also see Peretz, 1990 for right hemisphere dominance on contour processing). Also, Peretz and Babai (1992) reported that cerebral asymmetries in melody recognition depend on the tasks involved. When tasks required

listeners to recognize the overall pitch direction (contour) of a melody, the left ear (thus the right hemisphere) showed a quicker response than the right ear. On the other hand, when tasks required listeners to recognize a pitch change (interval) in part of a melody, the right ear (the left hemisphere) showed an advantage over the left ear. Therefore, Peretz and Babai suggested that perception of contour and that of intervals are two distinct types of processing for which different sides of the brain may be specialized. Peretz's notion of multi-level processing in melody perception is also supported by Schiavetto, Cortese, and Alain (1999) who, in their event-related potential (ERP) study, found distinct neural patterns for the processing of contour-violating and contour-preserving melodies.

### **Rationale for the present study**

Taken together, melody perception seems to involve extracting global properties as well as the processing of individual pitches. The present study concerns two issues. First, given that a melody is an auditory Gestalt object (that is, it is more than the sum of its ordered pitches), its global properties may help melody recognition even if the identities of the individual pitches are made uncertain. No studies have yet distorted melodies by means of degrading the pitches of the melodies. By degrading the pitches, the global quality of the melodies can be preserved while creating much uncertainty about the component units. If melody recognition employs holistic processing in the same way as in object and face recognition, and in speech perception, then melodies would still be recognized even if their individual pitches were ambiguous. In other words, if melody recognition involves extracting a global quality that is distinguished from the qualities of

individual pitches, then a melodic feature such as contour information might be accessible even without a clear perception of the identities of the individual pitches.

The second issue concerns the role of context in pitch identity. In vision and language (e.g., object superiority, configural superiority, word superiority effects), top-down processing often helps to perceive component units of a larger structure. In music, pitch judgment is found to be more accurate if the pitch is embedded in a series of pitches, creating a tonal context, rather than if a pitch is presented in isolation (see Dewar, Cuddy, & Mewhort, 1977; Warrier & Zatorre, 2002). Accordingly, it would also be possible that melodic context would help to disambiguate the degraded pitches of the present study.

The present study involved isolating the global properties of familiar melodies by diminishing the effects of feature-based identification. The method used in the present study was similar to the methods used in face perception studies that examined holistic processing of faces (e.g., Palmer, 1975; McKone, Martini, & Nakayama, 2001). In the face perception studies, individual features of faces were made uncertain by simplifying or degrading the features, or replacing them with other objects, so that only the global structure of the faces were preserved. Similarly, the present study minimized local cues to the identity of the melodies by reducing the sense of absolute pitch. This was accomplished by degrading the pitches with noise. Specifically, the tones of well-known melodies were replaced with bandpass-filtered noise bursts. This had the effect of degrading the pitch quality of the stimuli to such a degree that in the high bandwidth conditions, absolute pitch identifications were severely disrupted.

I first tested people's pitch identification abilities by presenting single degraded tones in isolation in order to verify that the stimuli were ambiguous or indefinite with respect to pitch. Second, I presented dyads composed of a pair of degraded tones in order



to ask the question of whether having more than one tone would enhance pitch perception through contextual cues. Finally, I presented listeners with melodies that consisted of these degraded pitches and tested their pitch identification abilities and their overall melody recognition. Note that I left rhythmic cues intact. However, in order to minimize the influence of rhythmic cues on melody recognition, I selected melodies that could not be identified on the basis of rhythmic cues alone, based on a previous study (Roberts & Levitin, 2001).

## METHOD

### *Participants*

Sixteen participants (6 men and 10 women) were recruited from McGill University. The age of the listeners ranged from 18 to 54 years ( $M = 23.94$ ,  $SD = 8.65$ ). All participants had more than ten years of musical training ( $M = 13.66$ ,  $SD = 3.13$ ), and had grown up in North America to ensure prior exposure to the melodic stimuli that we used. See Appendix A for ethics compliance.

### *Materials & Apparatus*

Stimuli were created using the computer program written in MATLAB (The MathWorks, 1998). The signals were obtained by filtering white noise with a biquadratic filter<sup>1</sup> (a second-order recursive bandpass filter) creating a spectrum with a single peak and a roughly symmetrical decay of intensity as the frequency deviated from this peak frequency in the higher and lower directions on a logarithmic scale. The sampling rate at which the stimuli were synthesized was 44100 Hz. The bandpass filters varied in their

---

<sup>1</sup> A bandpass filter is characterized by its central frequency,  $f_0$ , at which the power is maximum, and its two cutoff frequencies,  $f_1$  and  $f_2$ , at which the power has decayed by 3dB from the maximum. The central frequency,  $f_0$ , is defined as the geometrical mean of  $f_1$  and  $f_2$ . In other words, it is the square root of the product of  $f_1$  and  $f_2$ :  $f_0 = \sqrt{f_1 \times f_2}$ .

The bandwidth of the filter is the difference in Hertz between the two cutoff frequencies:  $BW = f_2 - f_1$ . For each condition, the bandwidth was specified in terms of a certain number of semitones, using the formula,  $r = f_2 / f_1 = 2^{n/12}$ , where  $n$  is number of semitones between the two cutoff frequencies  $f_1$  and  $f_2$ , and  $r$  is the ratio of their frequencies. We can then obtain the expression of  $f_1$  and  $f_2$  as a function of  $f_0$  and  $r$ :  $f_1 = f_0 / \sqrt{r}$  or  $f_2 = f_0 \sqrt{r}$ . For example, if a noise burst with central frequency of 440Hz and bandwidth of 4 semitones were to be created, the frequency ratio is  $r = 2^{4/12} = 1.260$ , and the cutoff frequencies would be  $f_1 = 440 / \sqrt{1.260} = 392.00\text{Hz}$  and  $f_2 = 440 \sqrt{1.260} = 493.88\text{Hz}$ . The frequencies 392.00Hz and 493.88Hz are equivalent to G4 and B4 respectively, which are in fact separated by 4 semitones. Note that the cutoff frequencies would be distance from the central frequency of the filter by exactly the same number of semitones in the higher and lower directions.

bandwidth as specified Table 1, and seven different bandwidths (where bandwidth indicates the total number of semitones by which the lower and upper cutoff frequencies are separated).

For the *Single-tone* condition, stimuli were single filtered noises whose centre frequencies ranged from A4 (440 Hz; the fourth A on the piano keyboard counting from the bottom) to G#5 (830.6 Hz; the fifth G# on the piano keyboard), for a total of 12 frequencies spanning a full chromatic octave. The six different bandwidths (see Table 1) were randomly assigned to the 12 different centre frequencies of the filtered noises. For example, one of the stimuli was a filtered noise whose bandwidth was one semitone-wide and whose centre frequency was C5. This method of random assignment was repeated in order to create 12 single filtered noises. Therefore, each bandwidth appeared with four different frequencies randomly selected for a total of 24 pitch x bandwidth combinations.

For the *Double-tone* condition, stimuli were 12 pairs of filtered noises with the same bandwidth per pair. The six different bandwidths (see Table 1) were randomly assigned to the 12 pairs. The centre frequency of the second noise of each pair ranged from A4 to G#5. Note that the random assignment of bandwidth to pitch was done separately for each condition. The interval between the center frequencies of the first and second noises was randomly assigned from 12 musical intervals (ranging from unison to major 7<sup>th</sup>). The centre frequency of each second noise was randomly assigned to one of the 12 centre frequencies, and the first tone was calculated using the randomly assigned interval size. For example, if the centre frequency of the second noise in a pair was A4 and was assigned the interval of perfect fourth, then the centre frequency of the first noise would be E4 (to create an ascending interval). Half of the intervals were ascending and the other half descending. Each tone was two seconds long and there was no pause

between the first and the second noise. The entire method was repeated in order to create 12 more pairs of filtered noises. For the *Interval* condition, exactly the same stimuli used in the Double-tone condition were used.

Table 1<sup>2</sup>

Bandwidths applied to filtered white noise

Bandwidth (semitone)	Context		
	Single-tone	Double-tone & Interval	Melody
0.5	√	√	
1	√	√	√
2	√		
3	√	√	√
6	√	√	√
9	√	√	√
12		√	√

For the *Melody* condition, twenty well-known melodies were chosen from Roberts and Levitin's (2001) study which examined melody recognition abilities when only rhythmic cues were given. In the study, rhythms of well-known melodies were clapped, and subjects were asked to identify the titles of each melody. It was found that some

<sup>2</sup> Note that the experiment was based on an unbalanced design (i.e., some BW conditions did not appear in all three contexts). BW0.5 and BW2 were not included in the design for the Melody condition because there were a limited number of available melodies that were unidentifiable by using only the rhythmic cues. BW12 for the Single-tone condition and BW2 for the Double-tone condition were excluded from the design for time management purposes (in order to reduce the number of trials). BW12 for the Single-tone condition was selected for exclusion based on the prediction that pitch identification performance would already reach a bottom effect at BW9. BW2 for the Double-tone condition was arbitrarily selected for exclusion in order to include BW12 in the condition. Later analyses show that this unbalanced design with missing conditions did not have any effect on the results (see Results).

melodies were identifiable using only rhythmic cues while other melodies were not. The selection for the present study was based on those melodies that were not identifiable by rhythmic cues alone. Examples of the chosen songs are “Rudolf the Red Noised Reindeer”, “Mary Had a Little Lamb”, and “Twinkle Twinkle Little Star” (see Appendix B for the full list). Each melody in the current study was assigned to one of the five bandwidths (Table 1). The centre frequency was assigned such that the last tone of each melody ranged from A4 to G#5. The average duration for the melodies was 14.53 seconds, and the average duration of the last tones of the melodies was 1.39 seconds.

The noise-stored digital audio files were synthesized prior to presentation, and were presented on a Macintosh computer using SoundApp software (Franke, 1993-2000) through a Harmon-Kardon amplifier, in a particular sequence for a particular subject. Listeners heard the stimuli diotically through AKG240 headphones. Sound pressure level was set to a comfortable level by each participant.

### *Procedure*

Participants were told that the experimental task consisted of identifying a tone that could range from A4 to G#5 on a keyboard (Casio CTK-100). Participants were also told that the stimuli they were going to hear might not sound very clear. They were allowed to hear the stimuli as many times as possible, and they could also hum or whistle in order to match the pitch of each tone. The stimuli were presented in a randomized order. For the Single-tone condition, listeners heard a single filtered noise burst and were instructed to identify the pitch by selecting a tone on the keyboard. For the Double-tone condition, listeners heard two noise bursts and were instructed to identify the pitch of the second burst by selecting a tone on the keyboard. For the Interval condition, listeners were instructed that they would hear the same stimuli as they heard in the Double-tone

condition, except that this time, they were verbally given the name of the tone corresponding to the pitch of the first noise burst of the two burst pair. Their task was to identify the pitch of the second burst of each pair on the keyboard given the first one. For the Melody condition, listeners were told that they would hear a melody made of fuzzy sounds, and that their task was to identify the last tone of the melody on the keyboard, and provide keywords or a title for each melody. If they knew the melody but were not able to give any verbal information, then they could also hum the melody. For the Double-tone, Interval, and Melody conditions, listeners were told that only the last pitches were in the range of A4 to G#5, and other pitches were possibly beyond this range. The order in which the three conditions were presented was randomized for each listener. When the experiment was completed, all of the melodies from the experiment were played in the form of clear tones and participants were again asked to identify them, in order to verify the familiarity of the tunes for each listener.

## RESULTS

### *Percentage of correct identification*

Table 2 is a summary of the average percentage of pitches that were correctly identified for each context across the different bandwidths (BW). Note that the BW conditions that did not have all three contexts satisfied – i.e., BW0.5-Single and Double, BW2-Single, BW12-Double and Melody context conditions – were excluded from this analysis. For this reason, in order to provide the ANOVA with an orthogonal design, BW0.5, 2, and 12 were excluded from the analysis. A 3 x 4 (Context x BW) repeated-measures analysis indicated that there was a main effect of bandwidth ( $F(3, 45) = 113.52$ ,  $p < .001$ ), showing that the percentage of correct pitch identification differed significantly across different bandwidths. However, the main effect of context was not significant ( $F(2, 30) = 0.67$ ,  $p = .52$ ), indicating that there was no difference between contexts in pitch identification accuracy. Also, there was no significant interaction effect between contexts and bandwidths ( $F(6, 90) = 1.96$ ,  $p = .08$ ).

However, in order to determine whether the excluded data weakened the analysis (i.e., whether the weakness of the context effect was due to the exclusion of the data), another repeated-measures analysis was performed with the data included. For this analysis, I “imputed” missing values for the four conditions that were not included in the experimental design (Single-BW12, Double-BW2, Melody-BW0.5, and Melody-BW2). For each subject, regression analysis was performed on existing scores of each subject in order to derive coefficients for a curvilinear function ( $y = c + bBW + aBW^2$ ), and then values for the missing BW conditions were imputed using this quadratic function equation. These imputed scores then made it possible to obtain a 3 x 7 ANOVA with an

orthogonal design. A repeated-measures analysis, using the imputed scores, yielded similar results to those found from the previous analysis (BW effect:  $F(6, 90) = 128.94$ ,  $p < .001$ ; context effect:  $F(2, 30) = 1.70$ ,  $p = .20$ ; interaction effect:  $F(12, 180) = 2.01$ ,  $p = 0.03$ ). Neither the observed significant main effect of bandwidth nor the non-significant effect of context was due to the data exclusion. The analysis yielded a significant interaction effect but the practical significance is unclear and needs to be explored in future studies with a balanced experimental design.

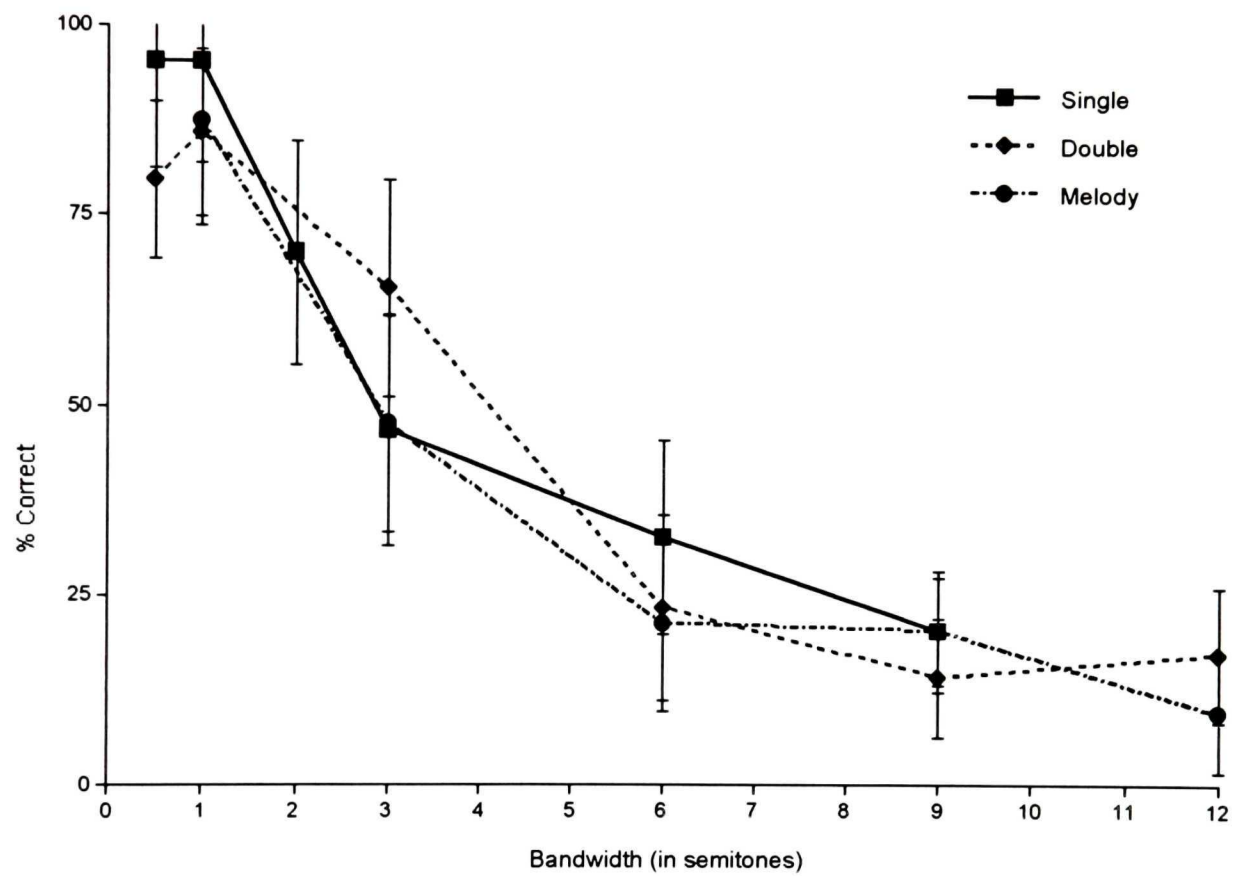
Table 2

Percentage correct for pitch identification

Bandwidth (BW)	Context		
	Single-tone	Double-tone	Melody
0.5	95.31	79.69	-----
1	95.31	85.94	87.50
2	70.31	-----	-----
3	46.88	65.63	47.88
6	32.81	23.44	21.31
9	20.31	14.06	20.31
12	-----	17.19	9.38

Figure 1 indicates the percentages of correctly identified pitches. Note that error bars for all the line graphs in our analysis indicate 95% confidence intervals. As shown, the overall trend is curvilinear. A complementary regression analysis was applied in order to describe the overall trend. The equations to describe the trend for the three contexts were  $y = 122.23 - 46.57BW + 3.76BW^2$  for the Single-tone,  $y = 60.05 - 12.67BW + 1.63BW^2$  for the Double-tone, and  $y = 72.69 - 11.40BW + 0.69BW^2$  for the Melody contexts.





*Figure 1.* Percentage of accurate pitch identification.

Melody recognition and pitch identification for the melody context (Figure 2) were also compared using a 2 x 5 (Task x BW) repeated-measures ANOVA. Note that here I compared data of different natures since the responses for the pitch identification were out of 12 possibilities whereas responses for the melody recognition were out of a virtually infinite number of possible melodies. However, the use of the ANOVA can be justified in that our test was thus a conservative comparison in the sense that it was biased in favour of finding that performance on the pitch identification tasks is better than on the melody identification task (i.e., it was biased against my research hypothesis). There was a significant main effect of context task ( $F(1, 15) = 137.67, p < .001$ ), indicating that listeners recognized melodies far better than they identified pitches. The main effect of bandwidth was also significant ( $F(4, 60) = 34.95, p < .001$ ), suggesting that listeners' performance differed across different bandwidths. There was a significant interaction between context and bandwidth ( $F(4, 60) = 9.78, p < .001$ ). As shown in Figure 2, the significant interaction suggests that the pattern observed in the figure is a reliable one, i.e., that listeners' ability to identify pitches of the melodies declined with decreasing bandwidth while their melody recognition ability did not worsen. For example (as shown in Figure 2), both the pitch identification (in the Melody condition) and melody recognition were quite good at BW 1 (87.5% and 100% accurate respectively), but when the BW was 12 semitones wide, only the melody recognition remained high (73%) while the pitch identification was just above the chance level (8%).

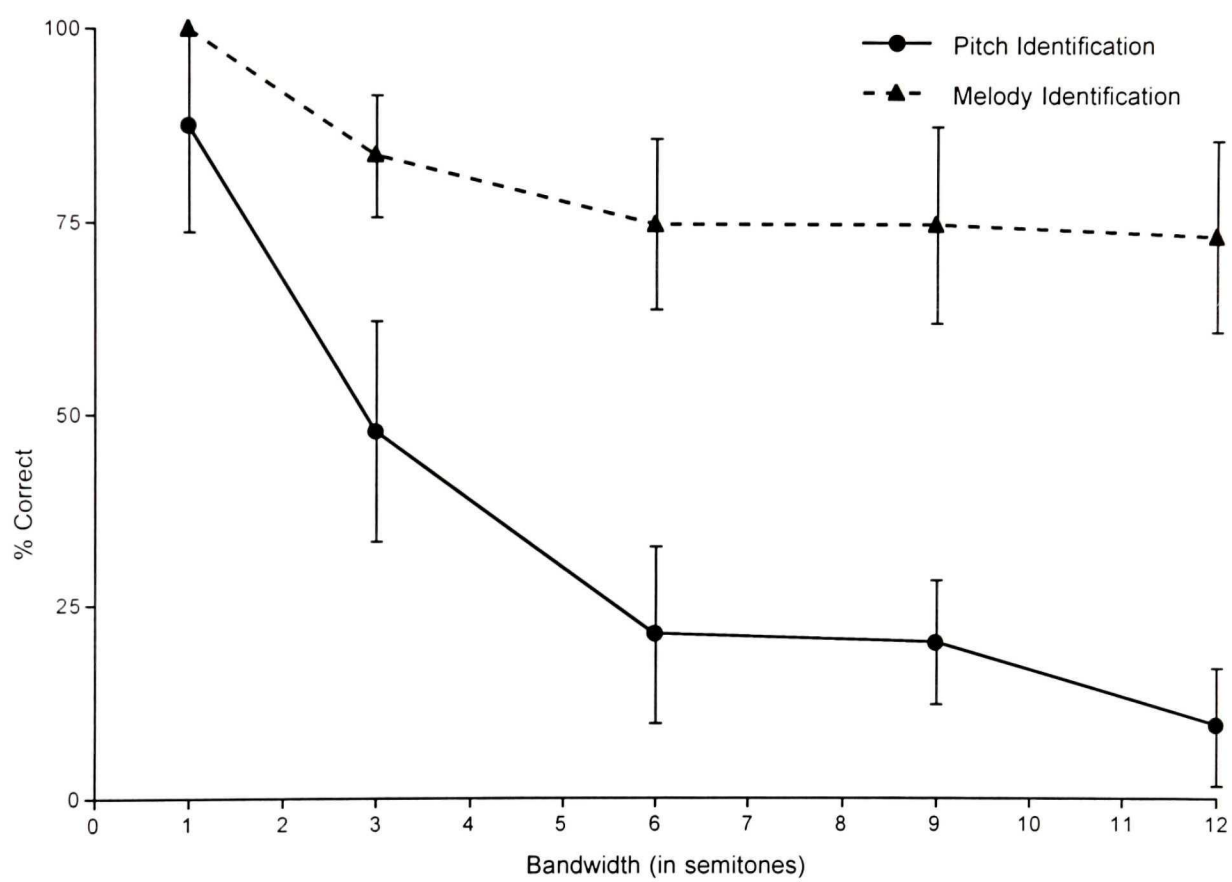
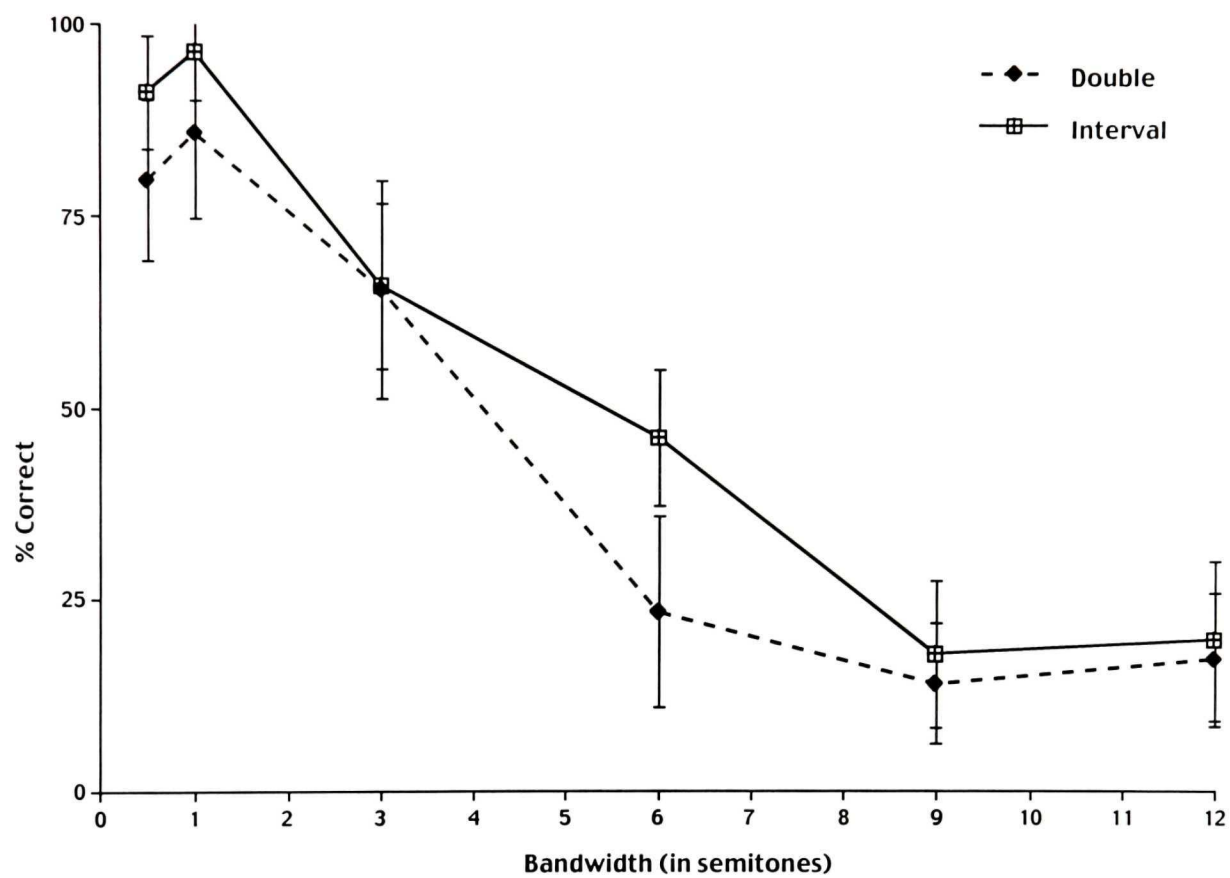


Figure 2. Percentage of accurate melody & pitch identifications in Melody condition.

In order to determine whether listeners had access to interval information, pitch identifications in the Double-tone ( $N = 16$ ) and Interval ( $N = 14$ ) conditions were compared (see Figure 3). A  $2 \times 6$  (Context  $\times$  BW) repeated-measures analysis indicated that there was a main effect of bandwidth ( $F(5, 65) = 49.28, p < .001$ ), showing that the percentage of correct pitch identification differed significantly across different bandwidths. There also was a main effect of context ( $F(1, 13) = 9.07, p = .01$ ), indicating that listeners were able to identify pitches of second tones better when the first tones were provided than when they were not. However, there was no significant interaction effect between contexts and bandwidths ( $F(5, 65) = 1.66, ns$ ).

In order to determine whether listeners had access to contour information, their sense of pitch direction (that is their judgment for a pitch going up or down) in the Interval context was analyzed. For example, if the first and second tones were C#5 and E5 respectively (ascending interval) and the identified second tone was D#5, the response was coded as having a correct pitch direction. However, if the identified second tone was lower than C#5, then the response was considered incorrect for pitch direction. The overall percentage of correctly identified pitch directions was 99.1% ( $N = 336$ ). The percentage of correctly identified pitch direction for the range of BW 0.5 to 6 was 99.1% ( $N = 224$ ), and the percentage for the BW 9 and 12 was 99.1% ( $N = 112$ ). The results suggest that listeners in general had a good sense of pitch direction when two tones were presented to them. Even when listeners' pitch identification was poor (i.e., at BW 9 and 12), their perception of pitch direction remained quite accurate.

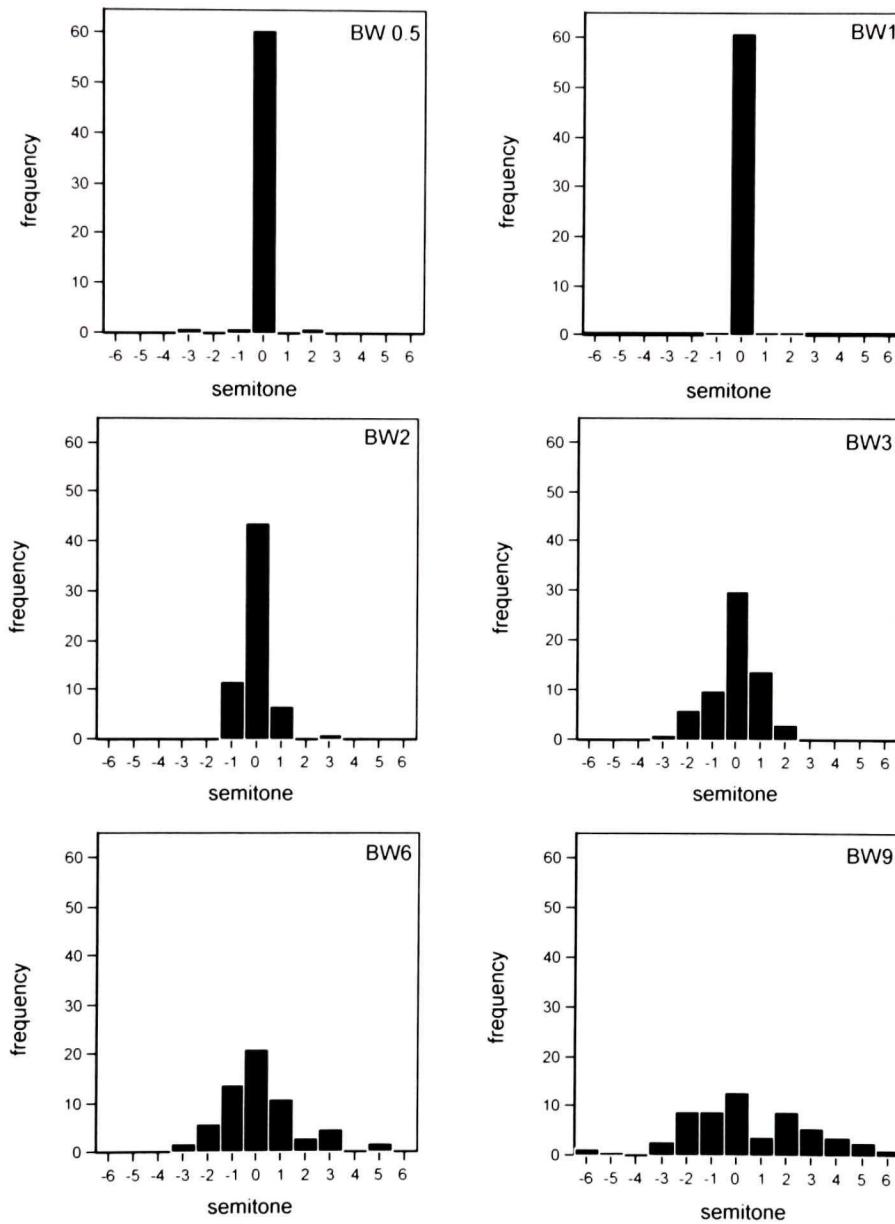


*Figure 3.* Percentage of accurate pitch identification in Double-tone & Interval conditions.

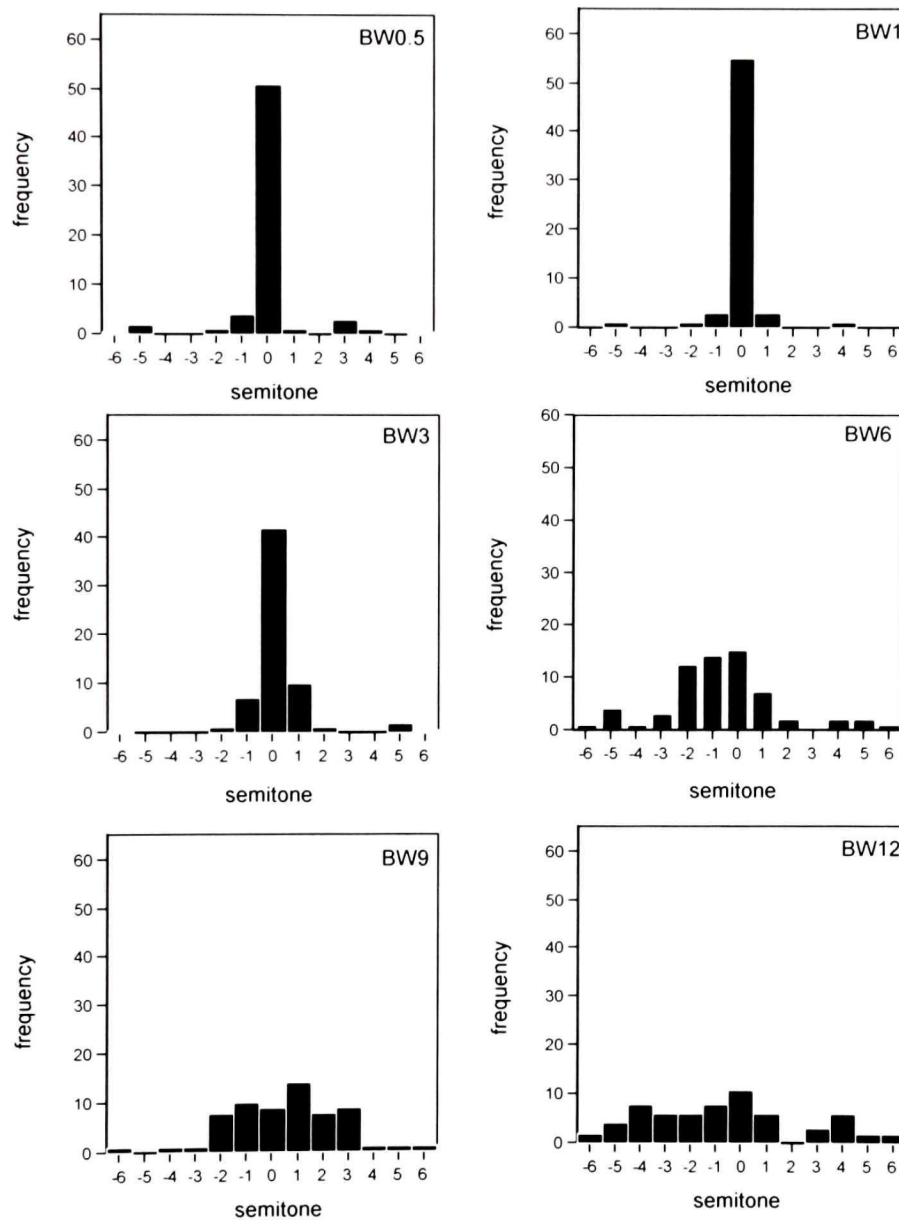
### *Semitone errors*

Listeners' responses were also analyzed in terms of semitone errors (number of semitones that responses deviated from the "correct" pitch – the center frequency of the filtered noise). Since the number of semitones between two pitches (ignoring which octave they are in) can be expressed in two ways, with the exception of pitches that are exactly 6 semitones apart, the number of semitones that was smaller was chosen. For example, the number of semitones between B and G can be either 8 or 4; so 4 would be chosen. This modulo 12 arithmetic is consistent with the notion of octave equivalence. Consequently, the maximum possible size of semitone error was 6. Figure 4 indicates semitone errors for each bandwidth in each context. Note that the negative/positive sign indicates the direction of the error, and the number of semitone errors of size 6 was divided equally into +6 and –6 scores. As shown in Figure 4, the error distributions were unimodal overall, with the mode at the correct pitch, confirming that listeners were in generally responding to the centre frequencies.

Table 3 summarizes the mean size of semitone errors for pitch identification for each context across the different bandwidths, ignoring the direction of errors. A 3 x 4 repeated-measures ANOVA revealed a main effect of bandwidth ( $F(3, 45) = 69.14, p < .001$ ); the main effect of context and the bandwidth-by-context interaction were not significant ( $F(2, 30) = 1.76, p = .19$ ;  $F(6, 90) = 1.66, p = .14$ ). These data suggest that with decreasing bandwidth, listeners tended to make larger semitone errors, but the context of the experiment did not have any influence on the errors. Note that this analysis excluded the data for the BW conditions that did not have all three contexts satisfied,

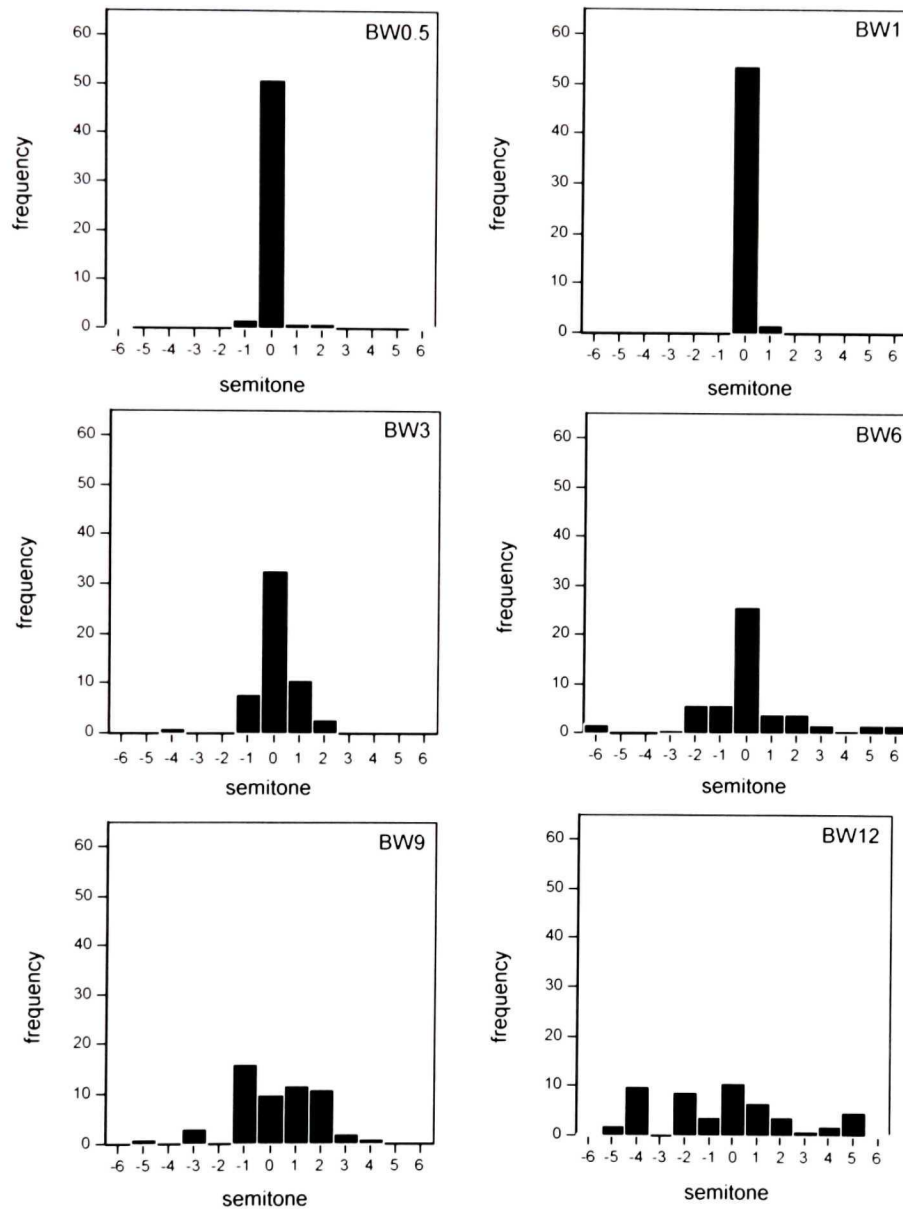


*Figure 4a.* Frequency distributions of errors of different magnitudes for Single-tone condition (in semitones).

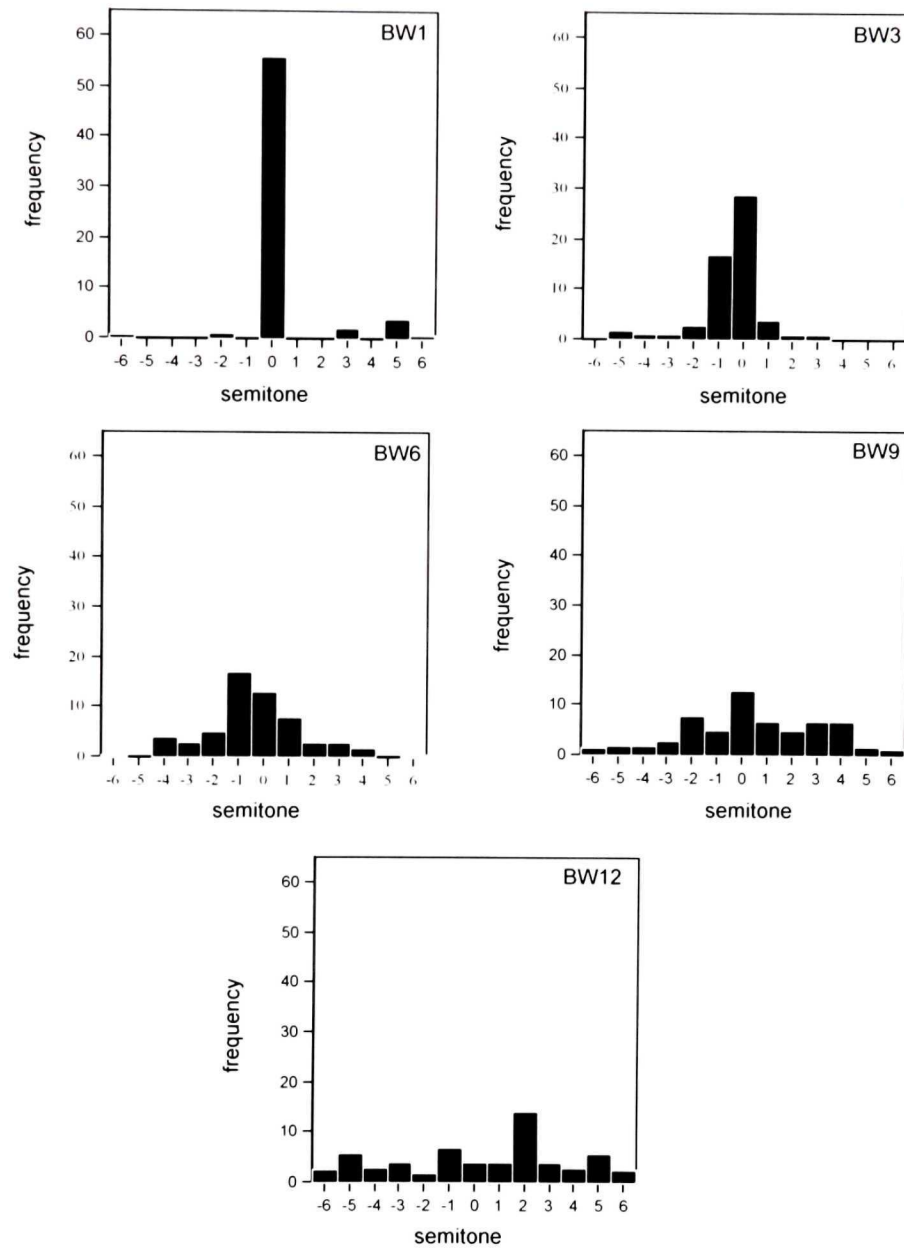


*Figure 4b.* Frequency distributions of errors of different magnitudes for Double-tone condition (in semitones).





*Figure 4c.* Frequency distributions of errors of different magnitudes for Interval condition (in semitones).



*Figure 4d.* Frequency distributions of errors of different magnitudes for Melody condition (in semitones).

similar to the exclusion in the analysis of the percentage of correct pitch identification. However, a  $3 \times 7$  repeated-measures analysis with the previously missing data included (using imputed scores for the missing cells) yielded similar results, but stronger results (BW effect:  $F(6, 90) = 90.43, p < .001$ ; context effect:  $F(2, 30) = 2.99, p = .07$ ; interaction effect:  $F(12, 180) = 1.64, p = .08$ ). They confirmed that the main effect of bandwidth was not due to the data exclusion, but gave “almost significant” probability values for context and interaction.

Table 3

Mean size of semitone errors for pitch identification

Bandwidth (BW)	Context		
	Single-tone	Double-tone	Melody
0.5	0.09	0.56	-----
1	0.06	0.27	0.53
2	0.34	-----	-----
3	0.70	0.58	0.88
6	1.19	1.75	1.53
9	2.03	1.73	2.22
12	-----	2.55	2.84

As shown in Figure 5, the overall trend for the semitone errors was linear as a function of bandwidth expressed in semitones, i.e., on a logarithmic scale. Therefore a constant multiplication of bandwidth yields a constant linear increase in the mean size of error. Using a regression analysis, I derived equations that describe the linear lines for each context. The equations can be expressed as  $y = .45BW$  for the Single-tone,  $y = .31BW$  for the Double-tone, and  $y = .30BW$  for the Melody contexts.

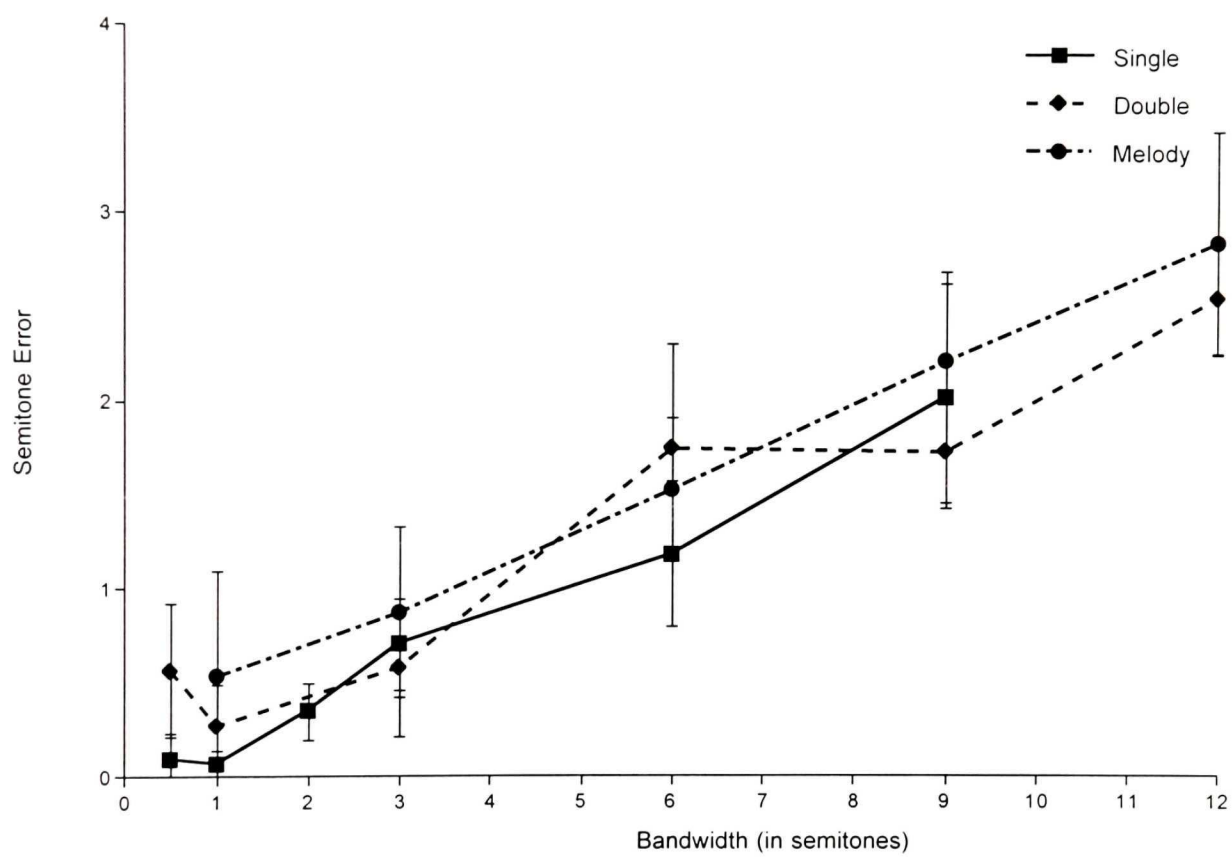


Figure 5. Mean size of semitone errors as a function of bandwidth.

## DISCUSSION

The finding that listeners' pitch perception weakened as the bandwidth of the filter increased regardless of conditions (Single-tone, Double-tone, Melody) implies that overall feature-by-feature identification (i.e., identification of the exact pitch) was disrupted by the change in bandwidth. For all of the conditions, listeners' responses were unimodal and symmetrical, and their most frequently perceived pitches were those of the center frequencies of the noise bands. However, as the bandwidth became wider, their pitch perception increasingly deteriorated. In the conditions with the widest bandwidth (BW12), listeners hardly had any sense of pitch, as measured by the ability to reproduce it on a keyboard. This implied that although listeners were responding correctly to the center frequencies at the lower bandwidths, they lost pitch information when there was too much noise. Thus, the bandwidth manipulation was indeed able to diminish the effect of precise information about the stimuli while preserving the global quality of the melodies.

The discussion will deal with the two issues that were raised in Introduction. My first question was as follows: given that a melody is a Gestalt object, is it possible to recognize the melody as a whole even if the component units of the melody are degraded? I will speculate as to the types of information that listeners were able to extract in order to recognize the distorted melodies. It is possible that listeners had access to contour information, or both contour and interval information. Another possibility is that listeners had partial access to interval information. That is, they might have been able to estimate approximate interval sizes but not exact sizes. Yet another possibility is that there could have been interaction between rhythmic and pitch cues so as allow recognition of the

identities of the melodies. The second question was as follows: does the melodic context help to identify the degraded pitches? I will discuss contextual influence on part identification in other domains of perception in comparison with the current study's contradictory finding that context did not aid the identification of individual pitches.

Regarding the first question, it was found that listeners were able to recognize melodies although they were not able to identify individual pitches that made up the melodies. Their response accuracy was still very high even when the bandwidth of the individual pitches became very wide, thereby disrupting their sense of pitch. This finding was more impressive in light of the fact that the melody task involved a free-choice recall test (i.e., it was not a forced-choice paradigm). Therefore, consistent with findings of holistic processing in other domains (object and face recognition, and speech perception), melody recognition seems to involve extracting a global quality that is distinguished from the exact qualities of individual pitches.

#### *Access to contour information*

Listener's melody identification was very good (73% accurate) even at the widest bandwidth although their pitch identification was very poor (9% accurate). Given that listeners must have relied on some features of the melodies to make the melody judgments, it is important to ask what those features might be. It is possible that contour information (the sequence of ups and downs regardless of pitch distance) was intact in the stimuli as the listeners generally had a very good sense of pitch direction (99% accurate contour judgment) in the Interval condition despite their poor pitch identification. This implies that even though listeners could not identify individual pitches of the melodies, they were usually able to judge whether a pitch in the melody was going up or down, thus

extracting the overall melodic contour pattern. However, note that in the study of Dowling and Fujitani (1970), recognition of familiar folk tunes based on contour information alone was rated at 59% accuracy. Although this performance was above chance (20%), it is far from being accurate considering that listeners were given one out of five melodies. Similarly, White (1960) found from his forced-choice task that when familiar melodies were distorted in such a way that relative interval sizes were distorted but contour information left intact, accuracy level for recognition was rated at 60% (10% was chance level). If the listeners in these studies were able to use contour information for recognizing familiar melodies, it seems that the performances should have been better than 59 or 60% correct. In comparison, the identification rate from the present study was found to be at 73% at the widest (and the most difficult) bandwidth – much higher performance than that obtained from the studies of White and of Dowling and Fujitani. One possibility for the higher performance in the present study could be that the two studies altered interval information by changing notes, as opposed to simply weakening the interval information in the present study. Their procedure changed the sizes of intervals. Hence it not only eliminated correct interval information but also introduced false interval information. Therefore, it is possible that the contour information was competing against the false interval information, thereby disrupting the melody recognition process. The present study might have disrupted interval information, but it is possible that it did so without introducing false interval information. It could be that listeners in the present study were able to employ pure contour information for recognition without the interference from false interval information, thereby performing better than the listeners in the other studies.

Yet another possibility for the performance gap could be that despite the accurate pitch direction judgment in the present study, contour information might not have been enough for listeners to trigger their melody recognition. Dowling (1994) and Edworthy (1985) argued that contour is salient in novel melodies and helps recognition, but that it does not play as important a role in the recognition of familiar melodies. They suggested that interval information is more useful in triggering memory for familiar melodies. Thus, it is possible that in the present study, there was information other than contour that aided our listeners in their melody identification, as described below.

#### *Access to interval information*

It is possible that listeners had partial access to interval information. The comparison of pitch identification in the Double-tone and the Interval conditions suggests that listeners were able to identify pitches of the second tones of dyads better when they were given the pitches of the first tones of the pairs than when they were not. This suggests that when they were not provided with these pitches, listeners might have had some interval information but they did not necessarily know the first pitches of the noise-burst pairs which would have enabled them to provide the correct second tones. When they were provided with the pitches of the first sounds, they were able to identify those of the second sounds using interval information. However, Figure 3 shows that although pitch identification was consistently better when the pitches of the first sounds were given to listeners than when no information was given, it was not much better at wide bandwidths (BW9 & BW12) where pitch perception deteriorated dramatically. Considering that the chance level was at 8%, listeners' performances were not much different from chance at wide bandwidths regardless of conditions (18% correct for the



Interval, 13% correct for the Double-tone condition at BW9, and 20% correct for the Interval, 16% correct for the Double-tone condition at BW 12). Also for both of the conditions, the lower limits of the 95% confidence intervals for the percentage of correct pitch identification were not far from chance level. Therefore, I am not convinced that people had good interval judgments when pitches were severely degraded. Listeners' melody identification might not have been based on good interval judgments. If not contour alone, and not intervals, then what were listeners relying their melody identifications on?

#### *Approximation of interval sizes*

It seems that listeners might have extracted more detailed information than contour but less precise information than exact interval sizes. It is possible that extracting the exact interval sizes is not necessary in order to identify melodies. White (1960), and Dowling and Fujitani (1970) suggested that listeners could rely on relative magnitudes of intervals in recognizing familiar melodies, thus absolute interval sizes were not necessarily required. Instead of extracting precise interval information, listeners in the present study could have approximated the degree of change in ups and downs of the pitches of the degraded tones in melodies. In other words, listeners might have had a rough idea of how much a pitch was going up or down in the distorted melodies.

#### *Interaction between different elements*

Another possibility for melody recognition is that different elements of the melody interact with one another. For example, it is possible that rhythmic cues and pitch cues interact to give rise to the identities of the melodies. The rhythmic cues alone were not sufficient to trigger listeners' memory for the particular melodies I used in the present

study because I chose only the melodies that were not identifiable by rhythmic patterns alone (Roberts & Levitin, 2001). However, it is possible that combination of rhythmic and pitch cues could yield more than additive effects. The rhythmic pattern alone could be weak information to trigger melody recognition, but when paired with pitch cues, rhythm could have become a stronger cue. It is possible that when rhythmic and pitch cues are present together, certain properties that did not exist before might emerge, and the emergent properties might aid melody recognition. Pomerantz et al. (1977), in discussing the configural superiority effect, (the orientations of lines were more discriminable when they were placed in a certain context rather than when they were presented without any context), argued that the effect worked in certain contexts because the lines interacted with other parts of the whole. They suggested that intersections of the lines with others to form an arrow or a right triangle create emergent properties that allowed changes in the diagonal line of the arrow and the hypotenuse of the triangle to be recognized more easily than the equivalent changes in the orientations of the lines when they are presented in isolation. Similarly, it is possible that the rhythmic cues were not sufficient to recognize the melodies, but when they are presented with degraded pitches, the combination might create an emergent property that allows for the melodies to be recognized.

Jones and her colleagues (Jones & Ralston, 1991; Jones, 1993; Jones & Pfordresher, 1997) proposed a framework called “joint accent structure”. Jones suggests that there exist two types of accents in musical patterns. Melodic accents mark changes that occur in pitch patterns in the melodies. Changes in a melodic pattern include pitch, interval, and contour change. Temporal accents mark changes or emphases in rhythmic patterns. Jones further suggests that when a rhythmic pattern and a melodic pattern are

coupled, the combination yields a global property called a “joint accent structure” in which a strong joint accent is created if a melodic and a temporal accent coincide at certain points in time, and a weak joint accent is created if only one of the accent exists at other points in time. The joint structure would contain different accents than what the melodic structure or the temporal structure alone would have contained, and this joint structure may be one of the things that we might perceive in melodies. Therefore, it is possible that listeners in the present study might have relied on the joint accent structure. However, the present study was not designed to test for interaction effects between rhythmic and pitch cues, so the results cannot confirm such an interaction. In order to find out, a follow-up study should test whether the distorted melodies can still be recognized when rhythmic cues are removed (e.g., play the melodies at an even beat, breaking up long notes into two or more single ones). If the removal of the rhythmic cues does not change the superiority of melody recognition performance over interval recognition, then one would be able to infer that there was no interaction between rhythmic and pitch cues and that the highly accurate melody recognition in the present study was not due to keeping the rhythmic cues intact in the melodies.

### ***Did context improve pitch perception?***

My second question was whether context would facilitate perceiving the degraded pitches. I presented intervals and melodies consisting of filtered noises in order to provide richer contexts for the noise burst that was to be judged. However, there was no effect of context, suggesting that whether noise bursts were presented in isolation or in the context of intervals or melodies did not make a difference in listeners’ pitch identification. Regardless of context, listeners’ pitch perception seemed to have

deteriorated as the bandwidth increased. This seems contradictory to the many findings of contextual influence on part identification in other domains of perception. For example, the object superiority effect is a phenomenon in which people are able to identify a component unit more accurately if it is embedded in the form of an object rather than in isolation (Weisstein & Harris, 1974). The context of an object allows for the line embedded in the object to be detected easily. Another example is the word superiority effect which demonstrates that a letter is much more easily detected when it is embedded in a word rather than if it is embedded in a non-word, or if it is presented alone (Reicher, 1969; Wheeler, 1970). Also, the phonemic restoration demonstrates the contextual influence on the perception of a missing phoneme. Depending on the sentence given, listeners tended to hear a different phoneme (Warren & Warren, 1970). Contrary to these findings, the interval or melodic context did not help perceive the pitches of degraded tones better than single presented degraded tones.

Perhaps melody recognition is different from word recognition or face recognition where perceiving the whole inherently implies perceiving parts. In word recognition, for example, once the whole word is recognized, one is able to identify each letter. Recognizing the entire word allows the reader to implicitly know what the individual letters are. The same principle applies to the phonemic restoration. Knowledge of the sentence that contains the word with a missing phoneme enables the listener to deduce what the phoneme is. However, the same rule does not apply to melody recognition. Melody recognition does not seem to require absolute pitch identification. It could instead be that the melodic context helped the listener to perceive relations between tones (the intervals) better than the absolute identity of the single tones. As Costall (1985) commented, perhaps “it makes little sense for the listener to treat pitch as an entity in

itself since, typically, pitch constitutes merely the *medium* of meaningful structures, be they musical or otherwise” (p.192). Instead of pitch, perhaps the smallest perceptual unit in melody recognition is an interval. It is possible that melodic context in the present experiment helped to disambiguate the intervals of the melody. This may be similar to the configuration effect in vision in which line *orientation* is discriminated better when embedded in a certain configuration than in isolation. Here, the angle of the line is the smallest perceptual unit rather than the line itself. In other words, what the context aided was perceiving the orientation of the line in relation to the context rather than the line itself (e.g., the line as composed of a narrow region of colour occupying a certain regions of space). Thus, considering that intervals may be the smallest perceptual units in melody recognition, it is possible that the melodic context helped to perceive the ambiguous intervals.

Trainor, McDonald, and Alain (2002) suggested that there are neural circuits that encode interval information (the pitch-distance relation between two tones) in melody recognition independently of pitch information. However, the present study did not test for a contextual influence on interval perception. Future studies should address this issue. For example, an experiment could be carried out to see whether there was a difference in perceiving intervals (consisting of noise bursts) in isolation and intervals that were part of melodies (consisting of noise bursts). The present experiment does not provide evidence regarding the issue since it was designed to examine the role of pitch in melody recognition and the role of context in *pitch* perception. If future studies find that degraded intervals are perceived more accurately when embedded in degraded melodies as opposed to in isolation, then this can indicate that melodic contexts may help disambiguate the intervals. (It might be necessary to use novel melodies instead of

familiar melodies because it is possible that listeners identify intervals of familiar melodies from their long-term memory. In this case, interval identification would not be a result of perception; it would simply be extraction from memory. Novel melodies would prevent this problem since listeners could not retrieve them from memory.) Furthermore, if it is true that context influences the perception of component units of a whole in the auditory domain in the same way as it does in other sensory domains, then it can be suggested that intervals (rather than individual pitches) are the smallest perceptual units in melody recognition (since melody recognition does not seem to help the recognition of individual pitches).

In summary, the present experiment sought to find a case in the auditory domain where the analysis of local features is not required in identifying an auditory object. It used melodies composed of degraded pitches to achieve the goal, and found that the exact pitches of individual elements were not necessarily required in the recognition of well-known melodies. Consistent with findings in other domains such as vision and speech, the present study has demonstrated that a larger structure can be recognized without a full recognition of the identities of its component units. This is in line with the Gestalt notion that perception of a whole is greater than (or different from) the sum of the perception of the parts. Global processing may be necessary in our auditory system, as in the general sensory system, in order to interpret sensory stimuli when parts of the stimuli are missing or feature information is ambiguous.

## REFERENCES

- Allegretti, C. L., & Puglisi, J. T. (1982). Recognition of letters in words and nonwords. *The Journal of General Psychology*, 107, 139-148.
- Boothroyd, A., & Nitttrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America*, 84, 101-114.
- Chalikia, M. H., & Warren, R. M. (1991). Phonemic transformations: Mapping the illusory organization of steady-state vowel sequences. *Language & Speech*, 34, 109-143.
- Chalikia, M. H., & Warren, R. M. (1994). Spectral fissioning in phonemic transformations. *Perception & Psychophysics*, 55, 218-226.
- Chastain, G. (1982). Scanning, holistic encoding, and the word-superiority effect. *Memory & Cognition*, 10, 232-236.
- Costall, A. (1985). The relativity of absolute pitch. In P. Howell, I. Cross, & R. West (Eds.), *Musical Structure and Cognition* (pp.189-208). London: Academic Press.
- Day, R. S., & Wood, C. C. (1972). Mutual interference between two linguistic dimensions of the same stimuli. *The Journal of the Acoustical Society of America*, 52, 175.
- Delis, D. C., Robertson, L. D., & Efron, R. (1986). Hemispheric specialization of memory for visual hierarchical stimuli. *Neuropsychologia*, 24, 205-214.
- Dewar, K. M., Cuddy, L. L., & Mewhort, D. J. (1977). Recognition memory for single tones with and without context. *Journal of Experimental Psychology: Human Learning & Memory*, 3, 60-67.
- Dewitt, L. A., & Crowder, R. G. (1986). Recognition of novel melodies after brief delays. *Music Perception*, 3, 259-274.
- Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America*, 49, 524-531.
- Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85, 341-354.
- Dowling, W. J. (1982). Melodic information processing and its development. In D. Deutsch (Ed.), *The Psychology and Music* (pp. 413-429). New York: Academic Press.

- Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, 50, 305-313.
- Dowling, W. J. (1994). Melodic contour in hearing and remembering melodies. In R. Aiello, & J. Sloboda (Eds.), *Musical Perceptions* (pp.173-190). Oxford: Oxford University Press.
- Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, 1, 30-49.
- Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception*, 2, 375-388.
- Effetto Arcimboldo*. (1987). Milan: Bompiani.
- Ehrenfels, C. von (1890/1988). On Gestalt qualities. In B. Smith (Ed.), *Foundations of Gestalt Theory* (pp.82-117). Munich: Philosophia Verlag.
- Eimas, P. D., & Corbit, J. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4, 99-109.
- Franke, N. (1993-2000). SoundApp (Version 2.7.3) [Computer software].
- Ganong III, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110-125.
- Garner, W. R. (1981). The analysis of unanalyzed perceptions. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual Organization* (pp.119-137). Hillsdale, NJ: Lawrence Erlbaum.
- Homa, D., Haver, B., & Schwartz, T. (1976). Perceptibility of schematic face stimuli: Evidence for a perceptual Gestalt. *Memory & Cognition*, 4, 176-185.
- Humphreys, G. W., & Riddoch, M. J. (2001). The neuropsychology of visual object and space perception. In E. B. Goldstein (Ed.), *Blackwell Handbook of Perception* (pp.204-236). Malden, MA: Blackwell Publishers.
- Jacewicz, M. M. (1979). Word context effects on letter recognition. *Perceptual and Motor Skills*, 48, 935-942.
- Johnson, N. F. (1975). On the function of letters in word identification: Some data and a preliminary model. *Journal of Verbal Learning & Verbal Behavior*, 14, 17-29.
- Johnson, N. F., & Marmurek, H. H. C. (1978). Identification of words and letters within words. *American Journal of Psychology*, 91, 401-415.



- Jones, M. R. (1993). Dynamics of musical patterns: How do melody and rhythm fit together? In T. J. Tighe, & W. J. Dowling (Eds.), *Psychology and Music: The Understanding of Melody and Rhythm* (pp.67-92). Hillsdale, NJ: Lawrence Erlbaum.
- Jones, M. R., & Pfordresher, P. Q. (1997). Tracking musical patterns using joint accent structure. *Canadian Journal of Experimental Psychology*, 51, 271-290.
- Jones, M. R., & Ralston, J. T. (1991). Some influences of accent structure on melody recognition. *Memory & Cognition*, 19, 8-20.
- Lassonde, M., Mottron, L., Peretz, I., Schiavetto, A., Hébert, S., & Décarie, J. (1999). Loss of global visual and auditory processing following right temporal lobe lesion. *Brain & Cognition*, 40, 162-166.
- Lawry, J. A., & LaBerge, D. (1981). Letter and word code interactions elicited by normally displayed words. *Perception & Psychophysics*, 30, 71-82.
- Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of Sixth International Congress of Phonetic Sciences, Prague, 1967*. Prague: Academia.
- McKone, E., Martini, P., & Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception & Performance*, 27, 573-599.
- Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience*, 9, 555-604.
- Palmer, S. E. (1975). Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp.279-307). San Francisco: Freeman.
- Palmer, S. E. (1980). What makes triangles point: Local and global effects in configurations of ambiguous triangles. *Cognitive Psychology*, 12, 285-305.
- Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. Cambridge, MA: The MIT Press.
- Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, 113, 1185-1205.
- Peretz, I., & Babai, M. (1992). The role of contour and intervals in the recognition of melody parts: Evidence from cerebral asymmetries in musicians. *Neuropsychologia*, 30, 277-292.

- Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 422-435.
- Purcell, D. G., & Stewart, D. G. (1988). The face-detection effect: Configuration enhances detection. *Perception & Psychophysics*, 43, 355-366.
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 275-280.
- Remez, R. E., & Rubin, P. E. (1984). On the perception of intonation from sinusoidal sentences. *Perception & Psychophysics*, 35, 429-440.
- Remez, R. E., & Rubin, P. E. (1990). On the perception of speech from time-varying acoustic information: Contributions of amplitude variation. *Perception & Psychophysics*, 48, 313-325.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, 212, 947-950.
- Roberts, M., & Levitin, D. (2001, August). *The role of rhythmic cues in melody identification*. Poster session presented at the meeting of the Society for Music Perception and Cognition Conference, Kingston, ON.
- Saumier, D., Arguin, M., & Lassonde, M. (2001). Prosopagnosia: A case study involving problems in processing configural information. *Brain & Cognition*, 46, 255-316.
- Schiavetto, A., Cortese, F., & Alain, C. (1999). Global and local processing of musical sequences: An event-related brain potential study. *Neuroreport*, 10, 2467-2472.
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, 46A, 225-245.
- The MathWorks, Inc. (1998). MATLAB (Version 5.2.1.1421) [Computer software].
- Tomiak, G. R., Mullennix, J. W., & Sawusch, J. R. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America*, 81, 755-764.
- Trainor, L. J., McDonald, K. L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience*, 14, 430-442.
- Versfeld, N. J., Daalder, L., Festen, J., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *The Journal of the Acoustical Society of America*, 107, 1671-1684.

- Warren, R. M. (1970). Perceptual restorations of missing speech sounds. *Science*, 167, 392-393.
- Warren, R. M., & Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American*, 223, 30-36.
- Warren, R. M., Bashford, J. A., & Gardner, D. A. (1990). Tweaking the lexicon: Organization of vowel sequences into words. *Perception & Psychophysics*, 47, 423-432.
- Warren, R. M., Healey, E. W., & Chalikia, M. H. (1996). The vowel-sequence illusion: Intrasubject stability and intersubject agreement of syllabic forms. *The Journal of the Acoustical Society of America*, 100, 2452-2461.
- Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics*, 64, 198-207.
- Weisstein, N., & Harris, C. S. (1974). Visual detection of line segments: An object-superiority effect. *Science*, 186, 752-755.
- Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, 1, 59-85.
- White, B. W. (1960). Recognition of distorted melodies. *American Journal of Psychology*, 73, 100-107.
- Wood, C. C., & Day, R. S. (1975). Failure of selective attention to phonetic segments in consonant-vowel syllables. *Perception & Psychophysics*, 17, 346-350.



## Appendix A

Faculty of Medicine  
3655 Drummond Street  
Montreal, QC H3G 1Y6  
Fax: (514) 398-3595

Faculté de médecine  
3655, rue Drummond  
Montreal, QC, H3G 1Y6  
Télécopieur: (514) 398-3595

CERTIFICATION OF ETHICAL ACCEPTABILITY FOR RESEARCH  
INVOLVING HUMAN SUBJECTS

The Faculty of Medicine Institutional Review Board consisting of:

LAWRENCE HUTCHISON, MD

SHARI BAUM, PHD

PATRICIA DOBKIN, PHD

HAROLD FRANK, MD

NEIL MACDONALD, MD

NANCY MAYO, PHD

WILSON MILLER, MD

LUCILLE PANET-RAYMOND, BA

HARVEY SIGMAN, MD

has examined the research project A04-B10-00 entitled "Development of the Theory of Auditory Scene Analysis Through the Creation of Critical Audio Example"

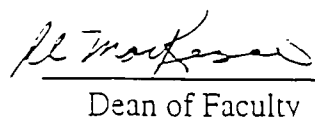
as proposed by: Dr. Daniel J. Levitin to Air Force Office of Scientific Research  
Applicant Granting Agency, if any

and consider the experimental procedures to be acceptable on ethical grounds for research involving human subjects.

April 17, 2000

Date

  
Chair, IRB

  
Dean of Faculty

Institutional Review Board Assurance Number: M-1458

## Appendix B

Melodies used for the Melody context and the familiarity test

---

Melody Title

---

Rock-a-Bye Baby

A Hard Day's Night

White Christmas

O Come All Ye Faithful

We Wish You a Merry Christmas

The Itsy Bitsy Spider

She'll be Comin' Round the Mountain

Hey Jude

If You're Happy and You Know it

Twinkle, Twinkle, Little Star

Hark the Herald Angels

Ode to Joy

Silent Night, Holy Night

Old MacDonald Had a Farm

For He's a Jolly Good Fellow

Pop Goes the Weasel

Frère Jacques

Mary Had a Little Lamb

Rudolf the Red Nosed Reindeer

O Canada

---