Pitch salience of various complex sounds

Citation for published version (APA):

DOI:
http://www.jstor.org/stable/40285262

Document status and date:
Published: 01/01/1984

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 02. Oct. 2023
Pitch Salience of Various Complex Sounds

ADRIANUS J. M. HOUTSMA
Institute for Perception Research (IPO)
Eindhoven
The Netherlands

Pitch salience of a variety of different complex sounds was measured through open-set melodic dictation tests using five musically experienced observers. The experimental task on each trial was to play back all notes of a four-note melody, randomly selected from an eight-note diatonic major scale, on an eight-note keyboard. Data were reduced to a correlation measure which addresses mostly the degree to which ordinal or contour information is preserved in the sequence of sensations, and also to a percent correct identification measure which tests preservation of ratio information. The two measures are in some cases very different, and it is proposed that those sounds that seem to convey mostly ordinal and little ratio information should not qualify as sounds that evoke true pitch sensations.

Introduction

Pitch is a very fundamental concept in music because music is essentially a variation in loudnesses, pitches, and timbres as a function of time. Most musicians, composers, and performers alike tend to treat pitch as a simple, well-understood attribute of sound without worrying too much about the various relationships between physical sound attributes and the subjective sensation of pitch. Almost all traditional musical instruments, including the human voice, are "special-purpose instruments" which the performer

1. Parts of this work were presented at the Symposium on Common Aspects of Processing of Linguistic and Musical Data, Tallinn, November 22–24, 1982 and are contained in the proceedings of this symposium.

Adrianus Houtsma carried out this work while he was lecturer in Music and lecturer in Electrical Engineering at the Massachusetts Institute of Technology. He is the author of several papers on pitch perception. He is currently Senior Research Associate at the Institute for Perception Research (IPO) at the Technical University of Eindhoven in The Netherlands.

Requests for reprints may be sent to Adrianus J. M. Houtsma, Institute for Perception Research, P. O. Box 513, 5600 MB Eindhoven, The Netherlands.

296
learns to play by developing a "feel" for the controls through trial and error, experience, and the help and encouragement of a good teacher. Understanding the physics of the instrument, that is, the various relationships between instrument controls and attributes of the acoustic output, is (fortunately) not a necessary part of learning to play an instrument, although it can often be of great help to a curious and intelligent player. Computer music, on the other hand, which involves a general-purpose instrument, puts a much larger burden on the musician who now must control (and therefore understand) all the important relationships between physical sound parameters and subjective sound attributes.

Almost every musician knows that the pitch of a sound has something to do with the rate of sound-pressure wave vibration. This idea was known to Galileo (1638) and became a fundamental concept in the hearing theory of Helmholtz (1863). Helmholtz thought that the ear analyzes a complex sound in real time into its Fourier components and that, in the case of a tonal sound, the fundamental component determines the subjective sensation of pitch. This idea, which was based on Ohm's acoustic law (Ohm, 1843), remained largely unchallenged for close to a century, but advances in modern psychoacoustics have shown that relationships between frequency and pitch are rather complex. The pitch of a pure sinusoidal tone, for instance, depends not only on its frequency but also on its intensity (Stevens, 1935). Complex tones, which are much more common in music than simple tones, evoke pitch sensations which are often determined exclusively by overtones (Seebeck, 1841; Schouten, 1938). Everyone has at one time or another heard this effect when listening to the bass section of a symphony orchestra through the kitchen table radio which transmitted only overtones of the cellos or double basses. Despite the obvious lack of high-fidelity in such music reproduction, there never is any confusion about melodic or harmonic information, individual pitches or octave ranges. Recent psychophysical experiments have revealed several other unusual pitch effects that have not found much use on the concert stage (yet) but are easily produced in an acoustic laboratory or sound studio. Some of these are "Huggins pitch" (Cramer and Huggins, 1958), "Sheppard pitch" (Sheppard, 1964), "binaural edge pitch" (Klein and Hartmann, 1981), repetition pitch (Bilsen and Ritsma, 1969), and negative after-image pitch (Zwicker, 1964). The issue addressed in this study is how one decides whether or not these effects discovered in the laboratory and earmarked as pitch effects are indeed pitch effects in the true musical sense.

The "official" definition of pitch as "... that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low" (ASA, 1960) appears much too broad because it addresses ordinal properties of sound. Percepts which musically would be classified as timbre, for example, qualities of dull versus sharp or dark versus bright,
easily fit within the boundaries of the definition. Musicians assign a much stricter meaning to pitch, as is evident in the use of specific harmonic and melodic intervals and tone scales, all of which imply ratio properties. A major third is different from a minor third not merely because it is a larger interval, but more specifically because both intervals correspond to specific frequency ratios (5/4 and 6/5, respectively). All sound sensations that have ratio properties have, of course, also ordinal properties, but the reverse is not necessarily true. Any useful definition of pitch should therefore be based on the ratio properties of the percept, and the question whether or not a perceptual effect is a true pitch effect can be answered by investigating to what extent these effects display ratio as opposed to merely ordinal properties.

This study describes a series of open-set melodic dictation experiments using a group of musically experienced subjects and employing a selection of sounds which, according to literature claims, have been reported to evoke pitch sensations. The selection is by no means complete and is intended only as an illustration of an experimental method. As controls, pure tones were used as sounds evoking clear and unambiguous pitch sensations, and 400-Hz wide bands of noise as sounds likely to evoke merely nontonal timbre sensations. Two methods of data analysis are presented which respectively test ordinal and ratio properties of the data, thus allowing a distinction between sounds evoking real pitch sensations and other that don't. The procedure, experimental results, and a general discussion are found in the following sections.

Experimental Procedures

The experimental task was comparable to a musical dictation. Subjects were presented sequences of four notes which were sampled randomly with replacement from a set of eight notes forming a diatonic major scale. They were required to play the note sequence back on an eight-note keyboard immediately following each presentation. No response-time limit was imposed. Four-note sequences were chosen because shorter sequences don't establish much of a melodic contour, whereas longer sequences are likely to involve short-term memory limitations (Deutsch, 1980). Our interest in this study is in perceptual qualities of sounds, and not in the subjects' ability to memorize sequences of such sounds. As a control, all subjects were first tested with pure-tone sounds and showed that they could identify series of four-note melodies at a level of 90% correct or better.

All note sequences had identical timing patterns, so that no rhythmic clues were available. Notes lasted 500 msec and were separated by 500-msec rests, except where indicated otherwise. The sounds used to represent the four notes of each sequence are listed below, and the category numbers
will be used for identification in the remainder of this paper. The octave range of the diatonic major scale is indicated for each sound.

1. Pure tones in the octave range of 400–800 Hz, used for control as sounds evoking clear and unambiguous pitch sensations.

2. Three-tone harmonic complexes of simultaneous successive harmonics of random order, presented diotically. The lowest harmonic number was random between 3 and 5, and the (missing) fundamental was chosen in the octave range of 400–800 Hz. Pitch sensations evoked by this kind of stimuli, often referred to as “residue pitches” in older literature, have been studied by many researchers during the past four decades (see de Boer, 1976 for a review).

3. Two-tone harmonic complexes of successive simultaneous harmonics presented dichotically. Fundamental and harmonic number ranges were the same as in number 2. This central pitch effect, which had important implications for the interpretation of “residue pitch,” was first reported by Houtsma and Goldstein (1971, 1972).

4. Three-tone harmonic sequences of successive diotic harmonics of random order (same as in number 2) with the three harmonics presented in rapid time-sequence from low to high for each note. Each harmonic lasted 40 msec, with 20-msec gaps in between, for a total of 160 msec duration of each of the four notes. Subjects’ ability to perceive pitches corresponding to the missing fundamental under these conditions was recently reported by Hall and Peters (1981).

5. Broadband noise multiplied with a pure tone in the octave range of 100–200 Hz. This yields a suppressed-carrier sinusoidally amplitude-modulated noise (SAM noise) with an envelope periodicity in the range of 200–400 Hz. Pitches evoked by periodic interruption or modulation of white noise have been reported, among others, by Miller and Taylor (1948), Pollack (1969), Burns and Viemeister (1976, 1981), and Houtsma, Wicke, and Ordubadi (1980).

6. Two 100-μsec pulses separated by a variable time interval between 2.5 and 5 msec. Bilsen and Ritsma (1969) reported that sounds with a superimposed echo delayed by 1–10 msec evoke a pitch corresponding to the inverse of the time delay. Note durations for these sounds are, of course, very short and melodies sound like four slightly tonal clicks separated by 500-msec rests.
### TABLE 1

**Number of Trials Presented for Each Subject and Stimulus Type**

<table>
<thead>
<tr>
<th>Stimulus Type</th>
<th>AH</th>
<th>RF</th>
<th>SW</th>
<th>LM</th>
<th>DH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>20</td>
<td>40</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>70</td>
<td>80</td>
<td>80</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>100</td>
<td>60</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>300</td>
<td>270</td>
<td>80</td>
<td>200</td>
<td>180</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>90</td>
<td>120</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>190</td>
<td>210</td>
<td>60</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>110</td>
<td>110</td>
<td>100</td>
<td>110</td>
<td>110</td>
</tr>
</tbody>
</table>

7. 400-Hz wide bands of noise with center frequencies in the octave range of 400–800 Hz. Signals were actually generated by multiplying 200-Hz lowpass noise (18 dB/octave rolloff) with a pure-tone carrier between 400 and 800 Hz. Such signals have periodic zero crossings but are perceptually indistinguishable from true bandpass noise. Although Fastl (1971), Bilsen (1977), and Klein and Hartmann (1981) have obtained rather consistent matches between pure tones and edges of noise bands, no one has ever claimed that such noise bands evoke tonal or pitch sensations.

All sounds were presented through earphones to individual subjects seated in a soundproofed chamber. No feedback was provided. Subjects were allowed, however, to interrupt a run of trials any time they wanted to listen to the diatonic scale played with the sound being tested. This was to reset their pitch reference in case it had drifted. The entire experiment was controlled by a computer which generated all random note sequences, controlled analog acoustic signal-generating equipment, and recorded responses. All subjects had some kind of formal musical training. AH, the author, is an amateur pianist and organist; RF, a professional music teacher and harpsichordist with absolute pitch; SW, an undergraduate science student who plays the cello in a major orchestra; and LM and DH are both music students (voice and piano/clarinet, respectively).

### Results

Raw data for each subject consisted of strings of eight numbers between one and eight, the first group of four designating the presented, the second group of four the perceived sequence of notes. Because each note represented a random choice out of eight possible notes, each note contained 3 bits of information for a total of 12 bits per trial. Table 1 shows the total
number of trials presented for each stimulus type to each subject. In general, more trials were devoted to stimuli that proved difficult, whereas fewer trials were spent on stimulus conditions that seemed to lead to near-perfect performance. Two different ways of treating the data are described next.

The first method involved the computation of a coefficient of correlation between the numbers representing the presented and perceived note sequences. All numbers representing presented note sequences were put in a single array forming a number set \([X]\) for each stimulus type and subject, with \([Y]\) being the set of corresponding response numbers. The correlation coefficient \(R\) is defined as:

\[
R = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2}}
\]

\(x_i = 1, 2, \ldots , 8 \)
\(y_i = 1, 2, \ldots , 8 \)

where \(x_i\) and \(y_i\) are the \(i\)th elements in \([X]\) and \([Y]\), \(\bar{x}\) and \(\bar{y}\) are the respective means of the elements in \([X]\) and \([Y]\), and \(N\) is the number of elements in each set. \(R\) is plotted in Figure 1 as a bar graph for all five subjects and all seven stimulus conditions. Values of \(R\) range by definition from +1 when
\[ y_i = x, \text{ for all } i, \text{ to } -1 \text{ when } y_i = -x, \text{ for all } i, \text{ and } R \text{ is zero when } x, \text{ and } y, \text{ are statistically independent. The correlation coefficient } R \text{ turns out to be a reasonably good measure of the ordinal relationship between } [X] \text{ and } [Y]. \]

One can easily see that if the condition \([x, \geq \bar{x} \text{ then } y, \geq \bar{y}]\) is true, every term in the numerator of Eq. (1) will make a positive contribution to \(R\). Therefore, if the perceived melodic sequence has the same general contour as the transmitted sequence, that is, if the numbers representing transmitted notes and perceived pitches are at all times in more or less the same place in relation to their respective means, the value of \(R\) will be relatively high. Correlation is very sensitive to melodic contour preservation, but surprisingly robust against failures to reproduce the precise magnitudes of distances or intervals between notes.

The second analysis method involved computation of the probabilities of the conditions “all four notes correct,” “three or more notes correct,” “two or more notes correct,” and “one or more notes correct” for each stimulus type and each subject. This measure is stricter than the correlation measure defined in Eq. (1); in fact, it is a very strict measure of ratio relations between transmitted and perceived note sequences. In order to identify a note correctly, one must have identified the previous note correctly and one has to identify the precise tone distance or ratio to the present note. Ratio properties imply ordinal properties, but the reverse is not true. A high identification score implies therefore a high degree of correlation but not vice versa. Correct identification functions for progressively stricter identification conditions are plotted in Figure 2 for all subjects and stimulus types. Each function represents only five points, which have been connected for convenience. Chance-level performance, that is, what one would get if all sound were turned off, is indicated in all panels by the dashed curve.

**Discussion**

The series of sounds used in this experiment is by no means complete. Many more sounds could have been tested in the same fashion. The purpose of this study was to address the question “what is pitch?” and to demonstrate an empirical method to answer that question. Furthermore it should be pointed out that the analysis methods used in this study for extracting ordinal and ratio properties from the data are probably not optimal. Ordinal properties could have been measured by counting the number of times transmitted and perceived intervals went in the same direction. Ratio properties could have been measured by tracking correct identification of successive intervals rather than notes. In the absence of a well-defined decision model for this kind of experiment, however, the choice between various ways to analyze the data remains somewhat arbitrary.

If one compares each subject's performance in Figures 1 and 2, one finds
that there are sounds whose perceptual correlates clearly display ratio properties and others that seem to have mostly ordinal properties. All stimuli that result in high note identification rates also show high correlation coefficients, which is of course implied, but there are also stimulus types which, for most subjects, show relatively high correlation coefficients combined with rather poor note identification performance. Sounds of this latter type, such as bandpass noise (number 7) or SAM noise (number 5) definitely evoke sensations that can be ordered on some kind of scale, and even allow one to measure just noticeable differences in center frequency or modulation frequency (Miller and Taylor, 1948). It has sometimes been argued (e.g.,
Hall and Peters, 1981) that when subjects can discriminate between two sounds on the basis of high versus low, there must be a pitch effect involved. Discrimination, however, requires only a sense of order. Therefore the position taken here is that only those sounds that pass the ratio test of correct note identification qualify as sounds evoking genuine musical pitch sensations.

Reviewing the results for the various sounds used individually, a number of observations can be made. All subjects perform almost perfectly with pure tones, which indicates that they were all sufficiently skilled to perform the complex task and that possible confusions in identification that may occur for other test sounds can be attributed to sensation errors and not to short-term memory or other central limitations.

Comparing performance for tone complexes of three successive random diotic harmonics (stimulus type number 2) to that of two successive random dichotic harmonics (stimulus type number 3) one observes that both of these stimuli evoke rather salient pitch sensations, but one also sees that the dichotic two-tone effect is weaker. One subject (DH) had considerably more difficulty with either sound type than any other subject. The fact that some listeners perceive a complex tone as a "Gestalt," having a unique pitch, whereas others perceive the same complex as several simultaneous partials, each with its own pitch, was already pointed out by Helmholtz (1863).

Stimulus type number 4, comprising three successive diotic and time-sequential harmonics, produced very interesting results. Hall and Peters (1981) reported that such stimuli evoked pitch sensations corresponding to the missing fundamental when the tones were presented at very low signal to noise ratios. This result was somewhat surprising given an earlier similar experiment (Houtsma and Goldstein, 1971) which turned out negative results on three subjects but was done at a considerably slower tone-sequence rate. Most of the trials in the present experiment were done at signal levels of 59 dB SPL in quiet, but half the trials of subjects AH and RF were taken in a broadband noise background of 68 dB SPL. Results under low S/N conditions were considerably worse than under signal in quiet conditions, which contrasts with the Hall and Peters findings. The data for subjects AH and RF in Figures 1 and 2 represent the averages of the high and low S/N conditions. Comparing results of simultaneous and time-sequential three-tone sequences one observes that for four of the five subjects ordinal recognition degrades noticeably and ratio recognition is reduced to slightly above chance level. This behavior does not support Hall and Peters's claim. Subject SW, however, was able to do quite well with sequential harmonics, although her performance is definitely not as good as with simultaneous harmonics. This sharp contrast between subjects' behavior poses an interesting puzzle. On the one hand behavior of the four subjects seems to point at a pitch percept that exists for simultaneous
harmonics but is absent when harmonics are presented one at a time. On the other hand, the performance of SW suggests that some kind of general pattern recognition scheme is used which is not very sensitive to simultaneous or nonsimultaneous presence of harmonics. Obviously, the behavior of SW deserves further study, particularly her ability to track a missing fundamental as a function of the rate at which harmonic signal components are sequentially presented.

Melody recognition with SAM noise signals (stimulus type number 5) and bandpass noise (stimulus type number 7) is rather poor in terms of individual note recognition, but quite strong in terms of contour recognition. Both kinds of sound apparently evoke sensations that can easily be ordered, as indicated by the high correlation coefficients, but are difficult to measure on a ratio scale. For 400-Hz wide bandpass noise this is not surprising because sensations of timbre can be ranked in an ordinal sense although the stimuli don't sound very tonal. Our SAM noise results contrast with those obtained by Burns and Viemeister (1981), who used a similar four-note dictation test on three musically trained subjects. They were able to get roughly 70% perfect recognition (all four notes correctly identified), against less than 10% for our subjects. In their experiments, however, subjects were told the starting note, an added-carrier signal was used resulting in a higher effective modulation index, and their envelope periodicity was in the 100–200 Hz range whereas ours was an octave higher.

For click pairs (stimulus type number 6) subject behavior was generally more variable than for most other stimuli. SW seems by far to have the most tonal percept, as indicated by the note identification function in Figure 2. Subjects RF and LM show identically poor interval recognition, but RF has much better contour recognition than LM. Single echoed clicks are about the weakest example of repetition pitch. The short duration of each click pair does not provide much redundant information, as is the case with echoed noise or echoed random pulse trains, for example, which may explain the relatively poor note recognition performance of most subjects.

Finally, if one compares Figures 1 and 2 for those stimulus types that seem to convey mostly ordinal information, one notices that note identification, poor as it may be, is still well above chance level as indicated by the dashed functions in Figure 2. Although ordinal properties do not imply ratio properties as mentioned earlier, it is important to keep in mind for this experiment that, if the stimuli in the note sequence provide ordinal information, significant clues are also being provided with respect to ratio identification. If for instance the second note is perceived to be higher than the first, all notes lower than the first note are eliminated as possible correct answers, thus reducing the range of notes to choose from. Although the exact statistics have not been computed, it is clear that chance level for correct note identification is considerably higher when ordinal information
is given on each trial. One should therefore not conclude that, because some sounds allow an observer to identify notes correctly at a level noticeably above chance level, the sound must therefore evoke some kind of pitch sensation.¹

References

¹ The endurance and patience of observers Rhona Freeman, Stephany Wingfield, Lisa MacCluskey, and David Horowitz are gratefully acknowledged. The author is indebted to L. D. Braida and H. S. Colburn for many helpful discussions. This research was supported by the National Institutes of Health (Grant NS11680–05).


