A psychophysical study of the perception of consonance and dissonance

STEPHEN METZ, ANNE D. PICK, and MARSHA G. UNZE
University of Minnesota, Minneapolis, Minnesota 55455

The purpose of this study was to assess how the qualities of consonance and dissonance are perceived. The participants first completed a pretest, in which they compared consonant and dissonant intervals with a standard interval. Then they completed a pair-comparison procedure, in which they made judgments of a series of pairs of intervals in terms of which interval sounded "better." The members of a pair were both consonant or both dissonant, or one was consonant and one was dissonant. Participants who had consistently identified intervals appropriately as either consonant or dissonant in the pretest perceived the consonant intervals as more or less equivalent. However, they perceived the dissonant intervals in terms of a dimension, with some such intervals being judged consistently as "better" than others. The order of subjects' preferences among dissonant intervals corresponded roughly, although not precisely, to the theoretical order described in some musicology literature and to the order generated by psychological judgments of music experts.

Combinations of musical tones that are judged to sound good or pleasant are called consonant, and combinations that are not are called dissonant. Pythagoras is credited with the observations that led to the knowledge that the frequency ratios of intervals judged consonant tend to be low or simple integer ratios, whereas the frequency ratios of dissonant intervals tend to be higher (Backus, 1969, p. 116). Several theories about the basis for consonance and dissonance have been proposed, the most well-known of which is attributed to Helmholtz. According to Helmholtz' theory, dissonance is perceived because of apparent beats that occur between overtones of the tones making up the combination or interval. The higher the integer values of the frequency ratios of the tones, the less the overtones will coincide, and the more apparent will be the beats that are said to be the basis for a perceived "roughness" in dissonant intervals (for extended discussions of theories of dissonance, see, e.g., Backus, 1969; Piston, 1978; Roederer, 1975).

Neither the beat theory nor any other theory accounts for judged consonance and dissonance, which are said also to involve "subjective" judgments, which themselves change over time and styles of music (Cazden, 1962). Van de Geer, Levelt, and Plomp (1962) assessed the meanings associated with perceived consonance and dissonance by asking a group of "intelligent" adults who were "all laymen where music is concerned" (p. 309) to rate a series of musical intervals according to Osgood's semantic differential categories (see Osgood,

Suci, & Tannenbaum, 1957, for a full discussion of this method). There were 10 scales in all, including such scales as high-low, sharp-round, active-passive, consonant-dissonant, wide-narrow, sounds like one tone/sounds like more tones, tense-quiet, rough-smooth. Applying factor-analytic strategies, the investigators identified three basic dimensions, in terms of which the intervals were judged: pitch, evaluation, and fusion. Evaluation includes the scales consonant-dissonant, euphonious-diseuphonious, and beautiful-ugly.

The factor of evaluation described by van de Geer et al. (1962) is related to definitions of consonance and dissonance in terms of frequency ratios of the component tones: In general, the more complex the tone ratio, the less is the evaluation of the interval. However, the relation is not perfect; the most evaluated intervals were the major and minor thirds and the minor sixth, whereas the octave, fifth, and fourth had the lowest integer ratios (2:1, 3:2, and 4:3, respectively). (Furthermore, in music theory, the fourth is itself sometimes classified as consonant, and sometimes as dissonant; see Castellini, 1962, p. 144; Piston, 1978, p. 7.) Thus, individuals' subjective preferences are not in perfect accord with objective and theoretical definitions of consonance and dissonance (see Cazden, 1962, pp. 211-212, for a discussion of consonance, the harmonic series, and theory of musical harmony). However, as Ward (1970) noted, it is not required that the two bases for judgment be brought into perfect agreement, or reconciled, as it were, but only that it be recognized that two criteria are applied to determine consonance and dissonance: one that is objective, and one that is subjective, preferential, and likely to change over time.

In 1914, Malmberg set out to determine empirically a rank order of musical intervals (within the octave bounded by middle C and the C above middle C) in

regard to degree of consonance. He worked in the laboratory of his mentor, Carl Seashore, and his goal was to provide a norm for constructing a test of individual differences in perception of consonance. The method he used was intriguing: He recruited a jury of eight experts—psychologists and musicians, some of whom were faculty members in the departments of psychology and music, including the head of the department of vocal music, as well as Professor Seashore. This jury then compared members of pairs of musical intervals as to their degree of consonance. Their procedure was as follows, in Malmberg's (1914, pp. 109-110) words: "The intervals were sounded simultaneously, with a duration of approximately two seconds, and were repeated as often as requested by any observer—sometimes as many as fifteen or twenty times. . . . Since it was necessary to arrive at one particular order which might be considered a norm, the observers adopted the plan of sitting together in the experiment, proceeding very slowly, and discussing all cases of doubt or difficulty, analyzing the situation, varying the conditions of stimulation, and refining observations. This proved a very great advantage since each of the eight observers, of different types of training, offered to one another criticisms and suggestions for points of view and in this way distinctions were developed and errors of observation were eradicated which might otherwise have passed unnoticed."

The resulting rank order was as follows (from most to least consonant): octave (C-C), fifth (C-G), major sixth (C-A), major third (C-E), fourth (C-F), minor sixth (C-A^b), minor third (C-E^b), diminished fifth (C-G^b), minor seventh (C-B^b), major second (C-D), major seventh (C-B), and minor second (C-D^b) (Malmberg, 1914, p. 120). This order, empirically derived from the judgments of a group of experts, corresponds reasonably well to the order derived from the theoretical and objective bases for defining consonance and dissonance (see, e.g., Cazden, 1959, Figure 4, p. 211).

Although there is a more or less specific acoustic basis for perceiving consonance and dissonance, there are demonstrable individual differences among ordinary adult listeners (i.e., nonexperts) in terms of the degree to which they do perceive some intervals as dissonant and others as consonant. In addition, we do not know that the distinctions made by individuals who do perceive consonance and dissonance correspond to the empirically and theoretically derived order of intervals. The description of Malmberg's (1914) procedure, for instance, in which it is noted that the judgment of some pairs of intervals was very difficult, suggests that some intervals may ordinarily be perceived as equivalent in terms of consonance rather than as differing in relative consonance. In other words, even though there is acoustic information that roughly specifies a dimension of consonance-dissonance, individuals' judgments of consonance or dissonance may not correspond precisely to that dimension. Thus, the purpose of this study was to determine the degree to which the empirically and

theoretically derived order of consonance-dissonance corresponds to how these qualities are ordinarily perceived, especially by individuals who do distinguish between consonant and dissonant intervals.

METHOD

Subjects
Fifty-one students in an introductory psychology course at the University of Minnesota participated in the study and received extra credit points. Individuals who were music majors or who had had considerable training in music or whose cultural backgrounds included non-Western music were excluded from participation.

Materials
The intervals were presented to the participants on audiotape via headphones at a comfortable listening level. The intervals were recorded onto the tape from a recently tuned grand piano at close range.

The eight intervals among which comparisons were made included the fifth (M5), major sixth (M6), major third (M3), fourth (4), minor seventh (m7), major second (M2), major seventh (M7), and minor second (m2). It will be noted that these eight intervals include the four most consonant intervals (excluding the octave) and the four most dissonant intervals in the order derived from Malmberg (1914). In addition, a minor third interval (m3), the middle value in the Malmberg order, was used for a pretest.

All intervals were within the octave whose lower bound was middle C. For the pretest, there were 16 pairs of intervals ordered randomly. They consisted of the m3 interval paired twice with each of the other eight intervals. Each interval was generated by selecting a note at random and constructing the interval from that note. The m3 interval always shared the lower note of the interval with which it was paired.

For a pair-comparison task that followed the pretest, there were 56 pairs of intervals ordered randomly. They were the 28 combinations of the eight target intervals paired with each other, and each pair was presented twice during the series. Each member of a pair was the first member for one presentation and the second member for the other presentation. The intervals for this task were constructed so as to be centered around the m2 interval, which was F-F#. Thus, for this task, the two members of an interval pair did not share the same lower note. One example of each interval was used to construct the pairs for this task.

For both the pretest and the pair-comparison task, the notes of an interval were played simultaneously, and the interval duration was approximately 3 sec. The intrapair interval was about 1.5 sec, and the interpair interval was 6 sec.

Procedure
Both the short pretest (approximately 4 min) and the longer pair-comparison task were completed by the participants in individual experimental sessions lasting approximately 30 min. The tasks were conducted in a quiet room with the experimenter present.

Each participant was informed about the nature of the task (i.e., that he or she would listen to pairs of musical intervals and would judge, for each pair, which interval sounded more pleasant. Prior to beginning their judgments, the participants listened to a few intervals so that they were acquainted with the types of sounds they would be hearing during the task. An intent of the instructions was to direct the participants' attention to the aspects of the sounds on which they should base their judgments. Thus "smooth," "clean," and "open" vs. "rough," "uneven," and "unpleasant" were among the adjectives used. For each interval pair, the participants recorded, on a score sheet, which interval sounded "better," or which they preferred according to the pleasantness criterion. The same instructions

were repeated for the pair-comparison task, which followed directly after the pretest. The participants were not informed of their pretest scores before proceeding to the pair-comparison task; it was implied to all participants that they were performing the task appropriately.

RESULTS

Pretest

The participants' responses for the pretest were scored as "correct" or "incorrect" in comparison with the ordinal dimension of consonance-dissonance derived by Malmberg (1914). Thus, judgments that the interval m3 was better than dissonant intervals m7, M2, M7, and m2 were scored as correct, and judgments that consonant intervals M5, M6, M3, and 4 were better than interval m3 were scored as correct. By these criteria, the participants made a total of 50% errors in judging the consonant intervals, compared with 6% errors in judging the dissonant intervals. This disproportionate number of errors for consonant intervals as compared with dissonant intervals was consistent across participants. Thus, for these participants, the m3 interval was much more like the consonant intervals and discrepant from the dissonant intervals than would be predicted from Malmberg's ordinal scale. It can be noted that the apparent position of the m3 interval in relation to the consonant intervals was more like what would be predicted from the frequency ratio theory of consonance and dissonance (see Cazden, 1959, p. 211, Figure 4) than it was like what would be predicted from Malmberg's order.

Pair-Comparison Task

The primary question to be answered by the results of this study is this: How are the qualities consonance and dissonance perceived by individuals who distinguish consonant from dissonant intervals? Thus, the most informative judgments are those of participants who made few, if any, errors in the pretest. Consequently, the participants were grouped according to their pretest scores, and the data for each group were analyzed separately. There were six participants who made either one or two errors in the pretest, and all of their

errors were in judging consonant intervals (i.e., on either one or two presentations of a consonant interval paired with the m3 intervals, they judged the m3 interval to be better). All four consonant intervals were represented among their errors.

The pair-comparison judgments for the group of six subjects were analyzed according to the procedures described by Guilford (1954, pp. 154-165) for deriving scale values from pair-comparison preference judgments. Extreme proportions are not used for this analysis (Guilford, 1954, p. 163), and because of consistent judgments across subject, it was not possible to generate a complete Z matrix. Therefore, calculations were made on an incomplete score matrix averaging across scale separations. The top row of Table 1 shows the Z score scale values derived from the analysis, with the musical intervals listed left to right in order of most to least judged dissonant. The scale values separating the intervals are an index of the consistency of the participants' judgments, and, of course, these are underestimates to the degree that they are based on judgments that were less than perfectly consistent across the six participants.

The first observation to make about the scale is that the comparative judgments of dissonant intervals were more consistent than were the judgments of consonant intervals. Both m2 and M7 intervals were consistently judged as more dissonant than M2 intervals, which, in turn, were perceived as more dissonant than m7 intervals.

The picture for consonant intervals is different. Consistent directional distinctions were not made among the consonant intervals, although all of the consonant intervals were perceived as "better" than the dissonant intervals. In fact, the degree of consistency in this regard is underestimated by the scale values in the table, since, as noted, extreme proportions were not used for deriving the scale. Eight of the 28 cells were unfilled because the six participants' judgments were in perfect agreement. They always judged intervals 4 and 6 to be better than intervals m2 and M7; they judged intervals M3 and 5 to be better than M7; they judged M6 to be better than M2; and they judged M3 to be better than m7.

In summary, the six individuals who initially consistently judged the dissonant intervals to be more dis-

Table 1

Scale Values (SV) for Eight Musical Intervals (I) for Subjects With Different Numbers of Pretest Errors

| Number of Pretest Errors | | | | | | | | | | | | | | | | | |
| 1-2 (n = 6) | | | 3 (n = 7) | | | 4 (n = 10) | | | 5 (n = 15) | | | 6 (n = 6) | | | 7-8 (n = 7) | | |
I	SV	D	I	SV	D	I	SV	D	I	SV	D	I	SV	D	I	SV	D
m2	.000	.000	m2	.000	.000	m2	.000	.000	m2	.000	.000	m2	.000	.000	m2	.000	.000
M7	.037	.037	M7	.157	.157	M7	.280	.280	M7	.264	.264	M2	.672	.672	M2	.265	.265
M2	.840	.803	M2	.606	.449	M2	.943	.663	M2	.589	.325	M7	.821	.149	M7	.265	.000
m7	1.107	.267	m7	.748	.142	m7	.951	.008	m7	.764	.175	m7	1.213	.392	m7	.617	.352
4	1.758	.651	4	2.129	1.381	4	1.881	.930	4	1.347	.583	M6	1.566	.353	M3	1.140	.523
M6	1.933	.175	M5	2.163	.034	M5	1.990	.109	M3	1.555	.208	4	1.877	.311	M6	1.226	.086
M3	2.077	.144	M3	2.235	.072	M3	2.195	.205	M6	1.563	.008	M3	1.928	.051	M5	1.301	.075
M5	2.121	.044	M6	2.298	.063	M6	2.411	.216	M5	1.678	.115	M5	2.051	.123	4	1.438	.137

Note—Intervals are listed in order of increasing judged consonance. D = difference from preceding scale value.

sonant than the standard, and the consonant intervals to be less dissonant than the standard, perceived differences among the dissonant intervals that were quite consistent with the theoretically and empirically derived ordering of these intervals. However, they did not so consistently perceive such differences among the consonant intervals presented in this study.

In subsequent rows of Table 1, the results of similar analyses of groups of subjects who made more errors in the pretest are depicted. Although their judgments were increasingly ambiguous in interpretability as their errors on the pretest increased, it can still be seen that the picture presented by their judgments mirrored that for the individuals who made only one or two errors on the pretest. The rank order of preferences for dissonant intervals was more consistent than that for the consonant intervals across these groups. In all groups except those who made the most pretest errors, the interval 4 was judged to be the least consonant of the consonant intervals, but the other three consonant intervals did not lie in any consistent order in relation to each other. However, the ordering of the dissonant intervals was maintained through all of the separate subject groups, except those who made six or more pretest errors. Finally, the distance separating the consonant from the dissonant intervals was typically the largest distance along the scale. This is consistent with the way the intervals were selected initially from Malmberg's (1914) derived order of intervals.

DISCUSSION

What was demonstrated in the present investigation are intraindividual differences in the degree to which there are consistent preferences among consonant intervals as compared with dissonant intervals, and interindividual differences in the consistency with which consonant intervals are distinguishable from dissonant intervals. In order to assess how these differences come about, we should want to know the degree to which listeners' judgments can be modified by experience, perhaps experience in which they listen closely to pairs of intervals or to specific intervals in their musical contexts. The development of such individual differences should also be important to investigate in order to better understand the nature and conditions of the perceptual learning that brings them about. In earlier pilot work, we found that after extensive training in making pair-comparison judgments of consonant and dissonant intervals, children 6-8 years of age showed improvement in distinguishing dissonant intervals from a standard, whereas they did not show such improvement in distinguishing consonant intervals from a standard.

In summary, the picture presented both by pretest judgments and by the pair-comparison judgments from this study is that adult listeners perceive dissonant intervals as varying in pleasantness, and the order of their preference corresponds quite well both to a theoretically and to an empirically derived order of dissonance. However, the same listeners perceive consonant intervals as more or less equivalent in pleasantness, even though there are acoustic differences that could be a basis for perceiving differences among them. The use of a grand piano instead of a computer or sine-wave generator to produce the musical intervals for this study was intentional, but it is an open and interesting question whether listeners perceive intervals produced by other instruments differently depending on the timbre of those instruments.

REFERENCES

BACKUS, J. The acoustical foundations of music. New York: Norton, 1969.
CASTELLINI, J. Rudiments of music. New York: Norton, 1962.
CAZDEN, N. Musical intervals and simple number ratios. Journal of Research in Music Education, 1959, 7, 197-220.
CAZDEN, N. Sensory theories of musical consonance. Journal of Aesthetics and Art Criticism, 1962, 20, 301-319.
GUILFORD, J. P. Psychometric methods (2nd ed.). New York: McGraw-Hill, 1954.
MALMBERG, C. F. The perception of consonance and dissonance. University of Iowa Studies in Psychology, 1914, 7, 93-133.
OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. The measurement of meaning. Urbana: University of Illinois Press, 1957.
PISTON, W. Harmony (4th ed.). New York: Norton, 1978.
ROEDERER, J. G. Introduction to the physics and psychophysics of music (2nd ed.). New York: Springer-Verlag, 1975.
VAN DE GEER, J. P., LEVELT, W. J. M., & PLOMP, R. The connotation of musical consonance. Acta Psychologica, 1962, 20, 308-319.
WARD, W. Musical perception. In J. V. Tobias (Ed.), Foundations of modern auditory theory (Vol. 1). New York: Academic Press, 1970.