DOCUMENT RESUME

ED 286 476 IR 012 809

AUTHOR	House, William J.; Davis, Cheryl
TITLE	The Recognition of Melodies by Humans and by Machine.
PUB DATE	24 Aug 86
NOTE	25p.; Paper presented at the Annual Convention of the American Psychological Association (94th, Washington, DC, August 22-26, 1986).
PUB TYPE	Information Analyses (070) -- Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Computer Simulation; Higher Education; Models; Multiple Regression Analysis; *Music; Predictive Validity; *Recognition (Psychology); Undergraduate Students
IDENTIFIERS	*Melody Recognition

ABSTRACT

The behavior of two real-time computer simulation models of melody recognition was compared with the performance of human subjects in this study. One of the models, INT1, recognized melodies by comparing specific intervals with stored intervals. The other model, CONT1, performed by comparing the contour of the stimulus melody with an array of melody intervals. The 20 intermediate and advanced psychology majors who volunteered to participate in the study were tested on the speed and accuracy of their recognition of 20 familiar melodies. In regression analyses, these data were regressed against the two computer models, and the results indicated that the contour analysis simulation (CONT1) was a more adequate predictor of human recognition than was the interval analyzer program. Four references are listed, and several graphs depicting the results are appended. (MES)

The Recognition of Melodies by Humans and By Machine

Research Reported at the 94th Annual Convention of the American Psychological Association

August 24, 1986

William J. House and Cheryl Davis

University of South Carolina Aiken

ABSTRACT

THIS EXPERIMENT WAS A TEST OF TWO COMPUTER MODELS OF MELODY RECOGNITION. THE BEHAVIOR OF TWO REAL-TIME SIMULATION PROGRAMS WAS COMPARED WITH THE PERFORMANCE OF HUMAN SUBJECTS IN A SPEED-ACCURACY FORMAT. ONE OF THE MODELS, INT1, RECOGNIZED MELODIES BY COMPARING SPECIFIC INTERVALS OF MELODIES WITH STORED INTERVALS; THE DISTINCTIVE FEATURE WAS SPECIFIC INTERVALS. THE OTHER MODEL PERFORMED BY COMPARING THE CONTOUR (SEQUENCE OF TONAL UPS AND DOWNS) OF THE STIMULUS MELODY WITH AN ARRAY OF MELODY INTERVALS.

IN REGRESSION ANALYSES THE HUMAN DATA WERE REGRESSED AGAINST THE TWO COMPUTER MODELS. THESE INDICATED THAT THE CONTOUR ANALYSIS SIMULATION WAS A MORE ADEQUATE PREDICTOR OF THE HUMAN RECOGNITION THAN WAS THE INTERVAL ANALYZER PROGRAM.

The Recognition of Melodies by Humans and By Machine

William J. House and Cheryl Davis

University of South Carolina Aiken

The fundamental question examined in this study was the nature of the matching process which must occur when a listener recognizes a melody which is being played. A relevant popular general model of pattern recognition is the distinctive-feature sort (see Estes, 1978). In these models, salient details of the stimulus are matched with a stored list of such details. If enough of these features are matched, recognition occurs. The object of the present study is to examine two possible distinctive-features of simple melodies and to ascertain which of these melodic features evokes more human-like pattern recognition in a computer model.

The critical feature of a melody cannot be the individual tones of the melody because human beings are capable of recognizing melodies no matter in what key the melody is played; the specific notes themselves are unimportant - rather, the distance between the notes (e.g. Plomp, Wagenaar, and Mimpen, 1973) or the relative simplicity of the ratio between the notes (e.g., House and Harm, 1979) is the crucial factor. Thus, the musical interval (e.g., octaves, perfect fifths, minor thirds, etc.), created by the relationship between the tones, is the factor which is invariant in the performance of a melody. If the musical interval is the critical feature by which recognition of melodies occurs, then a distinctive-feature system must match the intervals of the stimulus melody with some cognitive representation of the musical interval.

The first model in the present work is an implementation of a distinctive-feature matching system in which the musical interval itself is the critical melodic component by which recognition occurs. This proposition is embodied here in INT1, a computational interval-matching model of melody recognition.

Another possible salient psychological feature of melodies is melodic contour or the pattern of tonal ups and downs. Dowling (e.g., Dowling and Fujitani, 1971) has demonstrated that melodic contour may be the crucial means by which recognition of melodies happens. CONT1 is the present computational model of melody recognition which depends upon the matching of the abstraction of the musical intervals - their contour.

The Stimulus Representation

The "melodies" for both models were represented by a string of letters and other characters which were input from the keyboard of the computer as one might play a piano. Processing began with the introduction of the first "tone" letter. Subsequent letters were introduced in real time and corresponded to the note name of a given melody in conventional music; that is, letter names represented tones, a minus was a flat and a plus was a sharp. As apparently occurs in humans, these programs were designed such that the input melody sequence could be in any key but for purposes of the experiment only the keys of C major, Eb major, and F major were used. In order to simplify a very complex behavior, all of the melodies in this study were constructed with all of the tones having the same time value; that is, all melodies consisted of seven equal eighth-notes. Thus, the melodies were rhythmically identical.

Permanent Memory

Twenty melodies were selected on the basis of the variety of their structure, my judgment that they could be recognized in a steady eighth-note rhythm by human subjects, and my opinion that they were reasonably well known (see table 1). These melodies were stored as strings of musical intervals subscripted in a twenty by six (melody by interval) array. For example, the first six intervals of Yankee-Doodle were stored as

uni,ama2,ama2,dma3,ama3,dma2;

thus, a unison interval is followed by two ascending major seconds, a descending major third, an ascending major third, and a descending major second. The titles were stored in another array in such a manner that the coordinates for this title array corresponded with those of the melody array.

Immediate Memory and Partial Matching

Each tonal input resulted in the program fetching the tone's frequency (based on the equal tempered scale with A4=440 hz) and holding that value long enough to compute the ratio between it and the next tonal frequency:

$$R(j)=(FR[i])/(FR[i-1])$$

where j is one of the six possible ratios among the seven (i) tones.

A contingency table in which ranges of ratios were associated with musical intervals was next addressed. In the case of INT1 the exact musical interval was used as the basis of partial-matching. CONT1 assessed only the direction of movement (ascending versus descending intervals) for matching.

At the beginning of a test, all of the melodies had an equal probability of being the correct melody. Each melody was, thus, from the outset assigned the probability subscript of

$$P(m)=.05$$

where m is the number of a particular melody in memory. When a match occurred P(m) was increased additively

$$P(m)=[P(m)+.14].$$

When an interval (in the case of INT1) or the direction of movement (in CONT1) was obtained, it was compared with the appropriate interval in all of the melodies in the array of known melodies. If, for example, the first interval was a unison (wherein the first two tones are the same) then the program assigned all known melodies that also had an initial unison with an incrementing probability subscript. As the matching progressed, the melodies in memory began to acquire differing probability subscripts and groupings of structurally similar melodies evolved. When one of the melodies' probability subscript reached an arbitrarily set criterion, the program "guessed" that melody.

Feedback

After the program concluded that the partial input sequence was one or the other melody, the operator communicated with the program whether or not the response was correct. If the response was correct, the probability criterion for that melody in subsequent tests was decreased; otherwise the criterion was increased. Consequently, over trials, successes tended to cause the programs to "prefer" some melodies over others. Not surprisingly, I have observed humans employing this "anchoring" or favoring certain melodies in other melody guessing tasks.

Simulation Procedure

Each program was presented with five of the stored melodies (Yankee Doodle, Twinkle-Twinkle Little Star, My Bonnie Lies Over the Ocean, Danny Boy, and the principal theme from the Haydn Surprise Symphony) in a random order and in the several keys mentioned above over five trials. These particular melodies were chosen because they represented structurally different types of melodies. For example, Yankee Doodle and Twinkle are highly redundant in comparison to Danny Boy which has a relatively more entropic structure. The number of tones needed for the computer to make a guess was recorded along with both the number of correct responses and the melody name of incorrect responses. The results are presented below along with the results of the analogous human experiment.

The Experiment With Humans

Subjects

Twenty intermediate and advanced psychology majors volunteered to participate in this experiment. There were ten males and ten females and none of these people had any significant musical training or experience.

Stimuli

A tape was recorded with twenty melodies (table 1) all made up of seven sine-wave generated tones; each melody was played twice. This recording would be played for the subject prior to testing, demonstrating to her the pool of melodies from which she was to choose the correct ones during testing. Next, the same five melodies used in the simulation experiment above were randomly presented over five trials randomly in the keys of C, Eb, or F major. The sine-wave tones were generated on an Electrocomp 400 synthesizer with a Electrocomp 401 sequencer controlling the duration of the tone and the intertone interval. The tempo was set at a comfortable listening speed which was approximately two tones per second. They were recorded on an amplified Teac 2300S deck and were played back on Realistic brand stereo speakers.

Procedure

The subject was brought into a sound-tight laboratory room where she was placed in a comfortable chair facing the stereo speakers. The experimenter gave the subject a list of the twenty melodies typed on a sheet of paper and then proceeded to play the recorded seven tone melodies two times each pronouncing the name of the melody after it was played. The subject was allowed to repeat the presentation of the melodies if she wished.

The subject was told that she would be asked to listen to several of these melodies again and was to name the tune quickly - in as few tones as possible. Next, the subject was given five "dry-run" melodies to see if she understood the instructions. Then, the experimental trials began. The experimenter gave further instruction if either the responses were very slow or if too many errors were being made. The response as well as the tone on which the response was made were the dependent variables which were used for comparison with the computer behavior.

Results

Accuracy

Figure 1 is a graphic comparison of INT1's, CONT1's, and the human subjects' overall percentage of correct responses over five trials for each of the test melodies. Regression analyses were computed in which the human data were separately regressed against each of the computational models to assess each model's relative ability to predict the human subjects' behavior. CONT1 accounted for 40% of the human accuracy performance (R^2 =.40). The accuracy of INT1 had practically no correspondence to the human behavior (R^2=.008). Although not totally accurate in its recognition of the melodies, INT1 was much more accurate in "guessing" melodies than either CONT1 or the human subjects. INT1 was too good a recognizer to be an acceptable model of human recognition in terms of overall accuracy.

(A multiple regression was also computed using both of the models together as predictors. The human variance accounted for in this analysis was somewhat of an improvement over CONT1 alone (R^2 =.43). This opens for speculation the possibility of a future model in which the two approaches are somehow combined in preprocessing. This will be discussed more below.)

Speed

Figure 2 contains a comparison among the humans and the two computational models in their relative speed of recognition. The dependent variable in this case is the mean tone of the melody on which a correct recognition occurred over the five test trials on each of the five melodies. The humans were by far the slowest of the recognizers. This is because no particular care was taken in writing these programs to include an output component which would model human responding. When the recognition occurred in each of the programs, a print function simply produced the "guess" on the CRT. The response system in humans is clearly more complex and time consuming than that of the computer. Other experiments that we are doing now require subjects to press a button at the point of recognition and then make their oral response. In this way the computer's speed advantage is lessened. However, none of these matters influence the particular type of analysis being presented here.

Regression analyses of the mean-note-of-recognition were computed, as above, regressing the human data against each of the models. In contrast to the accuracy results described earlier, INT1 was a somewhat better model for the human recognition (R^2=.12) than CONT1 which accounted for virtually none of the variance (R^2 =.03) in mean-note-of-recognition. (As in the accuracy analyses, a multiple regression revealed the potential of an interaction between the models in modeling human recognition (R^2=.20).)

Learning

Figure 3 compares the models and the subjects on the mean percentage of correct recognitions on all melodies across the five trials. The human subjects demonstrated significant improvement over trials ($F[1,4]=44.31$, $p=.006$). Neither of the models' slopes demonstrated that learning had taken place ($F's<1$).

When the human accuracy data across trials was regressed against each of the models, INT1 accounted for 52% of the variance (R^2=.52) but the coefficient of correlation was negative! CONT1 only explained 2% of the experimental variance.

Insofar as increased speed of recognition over trials is a measure of learning (see figure 4), the humans produced evidence of improvement over trials (slope= -.29 F[1 4]=36.8, p<.01). Some trivial learning was observed in INT1 (F[1,4]=7.84, p=.07) but little confidence can be put in the negative slope created in the recognition by CONT1 (slope= -.08, F[1,4]=4.0, p=.14).

Regression analyses, however, demonstrated that both programs produced significantly accurate predictions of the human data with CONT1 (R^2=.72) doing much better modeling. INT1 accounted for 57% of the experimental variance (R^2=.57).

Confusion

Figure 5 displays the relative confusion found in the human recognizers as compared with the two models. For all of the times that the melody-name was guessed, the percentage of the time that the name was correct is presented. Thus 100% meant that the melody-name was never guessed correctly (total confusion). Neither of the computational models accurately predicted the human response; the human subjects were by far the least confused of the recognizers. CONT1 and INT1 accounted for 1 and 3 percent of the human variance respectively and both together only accounted for 3 percent.

Summary

Accuracy

CONT1 was the better model of human accuracy in recognition of these melodic patterns. INT1 was more accurate than either CONT1 or the human recognizers. This is not surprising when one considers that INT1 always has accurate although partial information from which to make decisions. Its only failures were the result of the effects of feedback and the arbitrary guessing criterion. CONT1, on the other hand, had to deal with the abstraction of melodic intervals, the sequence of ups and downs, which involved far less accurate information content than the exact musical intervals. Consequently, CONT1 was inaccurate - as were the human subjects. The inaccuracy of this model was not haphazard though; so much like the human behavior was CONT1's performance that this model predicted the human recognition very well (R^2=.40).

Speed

The modeling of the speed of recognition in humans by CONT1 and INT1 was unimpressive. I feel that much work needs to be done in simulating the reaction of humans in such decision making as was being considered in this study. Clearly, the humans were going through more processing than the computational models were. This modeling could be accomplished by separating the recognition stage from the reaction stage in an empirically justifiable manner. In the present pilot work there was no distinction between recognition and reaction in the two models. They were, thus, given an advantage over the human recognizers in the speed of recognition.

Learning

CONT1 modeled the human course of learning better than INT1 (R^2=.72). Learning was slight in humans but, indeed, appeared to be happening. INT1 was also a reasonable predictor of the human behavior (R^2 =.57) but this should not be surprising because both models had identical feedback mechanisms. Thus, the superiority of CONT1 is even more important; the difference in efficacy between the two models must be in the pattern recognition component rather than in the feedback.

Confusion

In a general sense, both of the present theoretical models of melodic pattern recogni ion were more accurate and faster than the human subjects. However, as difficult as the experimental task was for the humans, the humans were remarkably less confused than the models.

Conclusion

The stronger model of human pattern recognition in the present pilot work appears to be CONT1. Even with major failures (such as the total failure of CONT1 to contend with the Surprise Symphony Theme), when one of the models succeeds it is usually CONT1. However, when it fails, it is sometimes augmented in its ability by INT1; often together they account for more human performance than they do separately. The implication here is clear: the rapid recognition of melodic sequences probably involves multiple stages of analysis triggered by some stimulus characteristic perhaps in a preprocessing stage. In the case of CONT1 and INT1, one can conceive of a model combining the two recognition approaches such that a preliminary decision is made regarding the type of melody being presented (e.g. a folk melody versus a children's tune). This decision could be accomplished using a heuristic estimate of the entropy of the melody. The results of this branch of the program would lead to either a CONT1 or a INT1 type analysis.

References

Estes, W.K. Perceptual processing in letter recognition and reading. In E.C. Carterette & M.P. Friedman (Eds.), _Handbook of Perception_, Vol IX. New York: Academic Press, 1978.

Dowling, W. J. and Fujitani, D. Contour, interval, and pitch recognition in memory for melodies, _Journal of the Acoustical Society of America_, 1971, 49, 524-531.

House, W.J. and Harm, O.J. Sensitivity to the elements of pure tone auditory dyads. _Journal of Auditory Research_, 1979, 19, 111-116.

Plomp, R., Wagenaar, W.A., and Mimpen, A.N. Musical interval recognition with simultaneous tones. _Acustica_, 1973, 29, 101-109.

Table 1

Melodies in Permanent Memory

Yankee Doodle
Twinkle Twinkle Little Star
Old McDonald
Haydn Surprise Symphony Theme
Auld Lang Syne
Swanee River
My Bonnie Lies Over the Ocean
I've Been Working on the Railroad
Camptown Races
Comin' Through the Rye
Flow Gently Sweet Afton
Drink to Me Only With Thine Eyes
Danny Boy
Dvorak New World Symphony Theme
Oh Susanna
Aloha Oe
Three Blind Mice
Red River Valley
Dixie
Frere Jacque

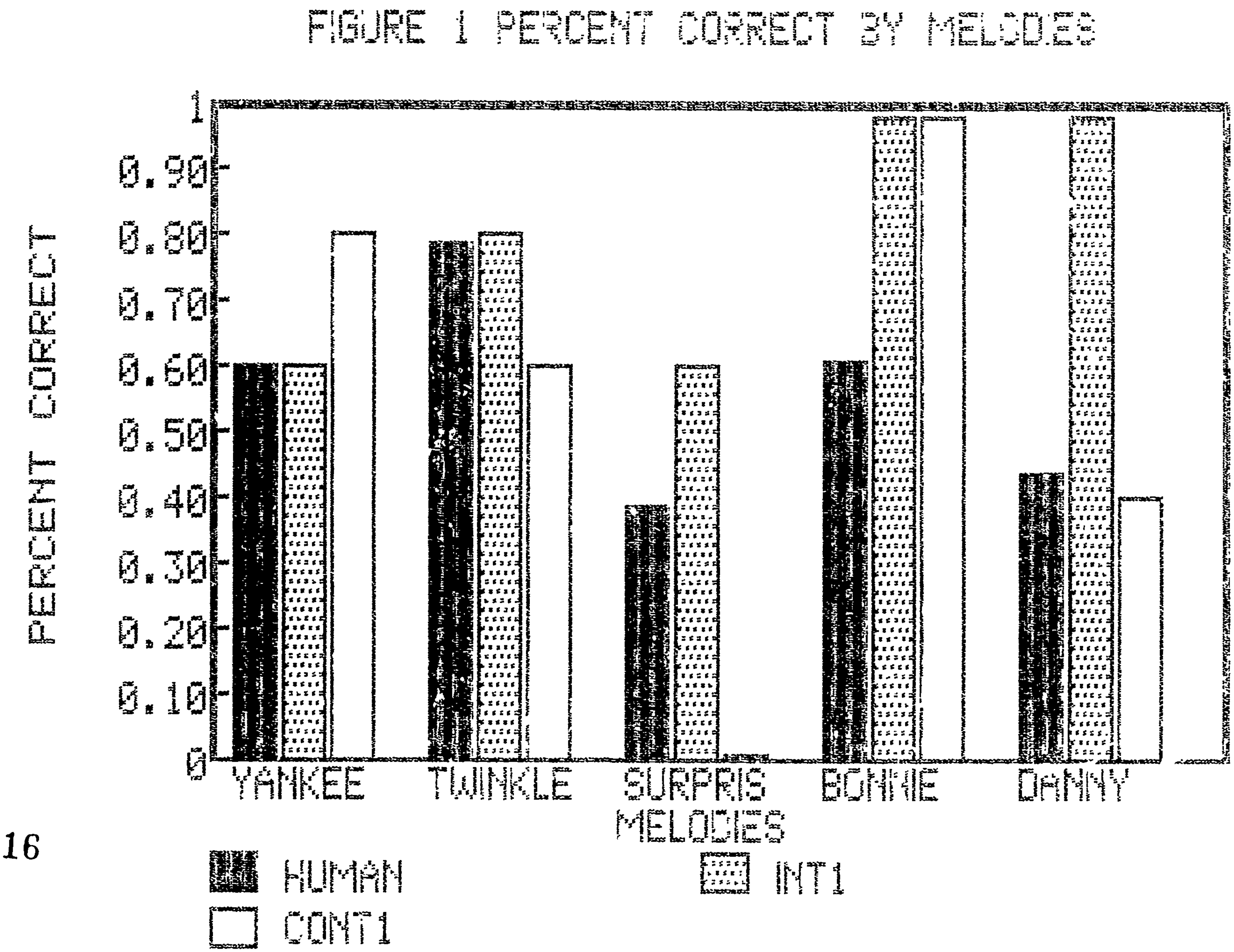
FIGURE 1 PERCENT CORRECT BY MELODIES
PERCENT CORRECT
1
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0
YANKEE
TWINKLE
SURPRIS
BONNIE
DANNY
MELODIES
HUMAN
INT1
CONT1

FIGURE 2 MEAN NOTE CORRECT RECOGNITION

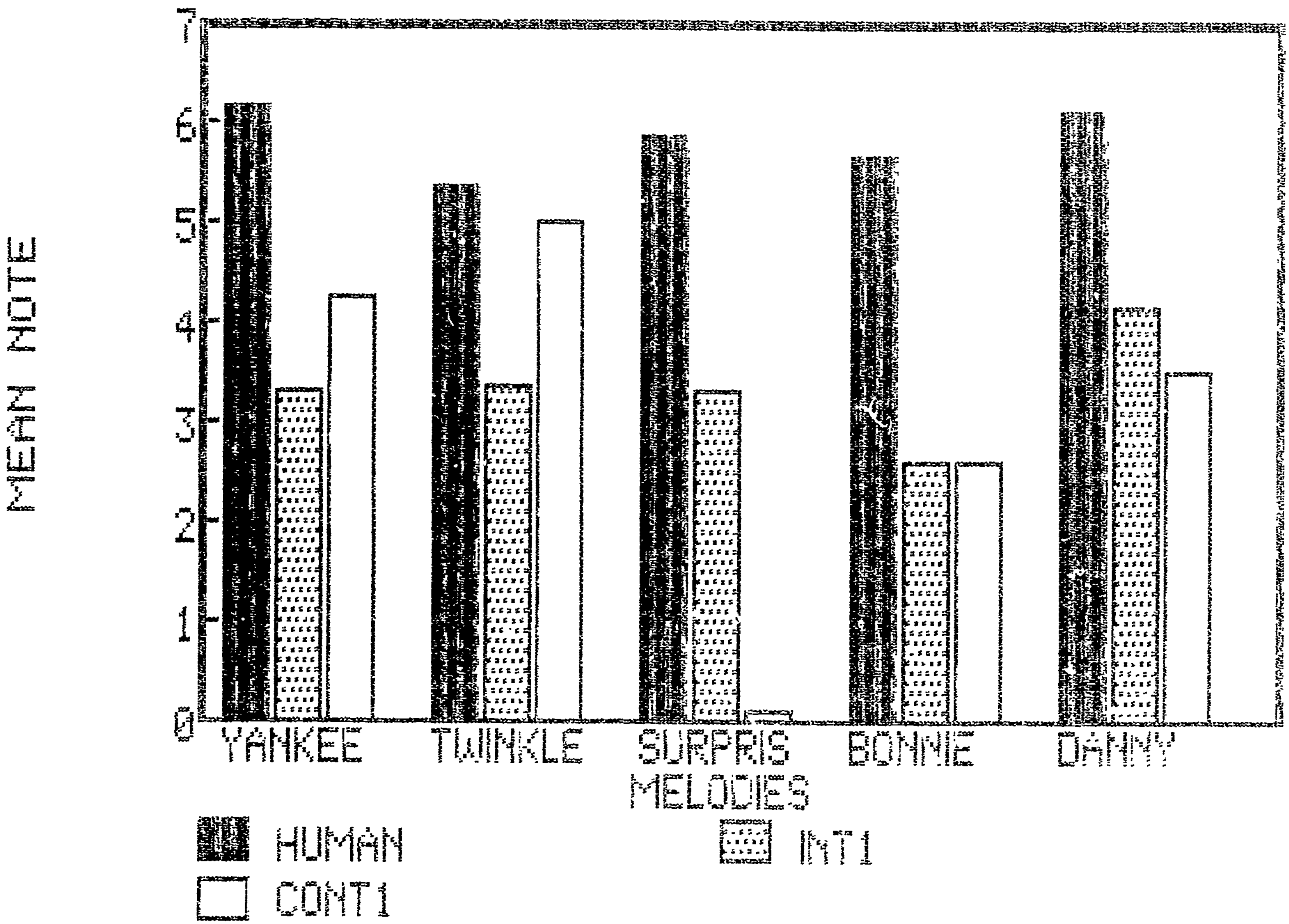

FIGURE 3 PERCENT CORRECT BY TRIAL

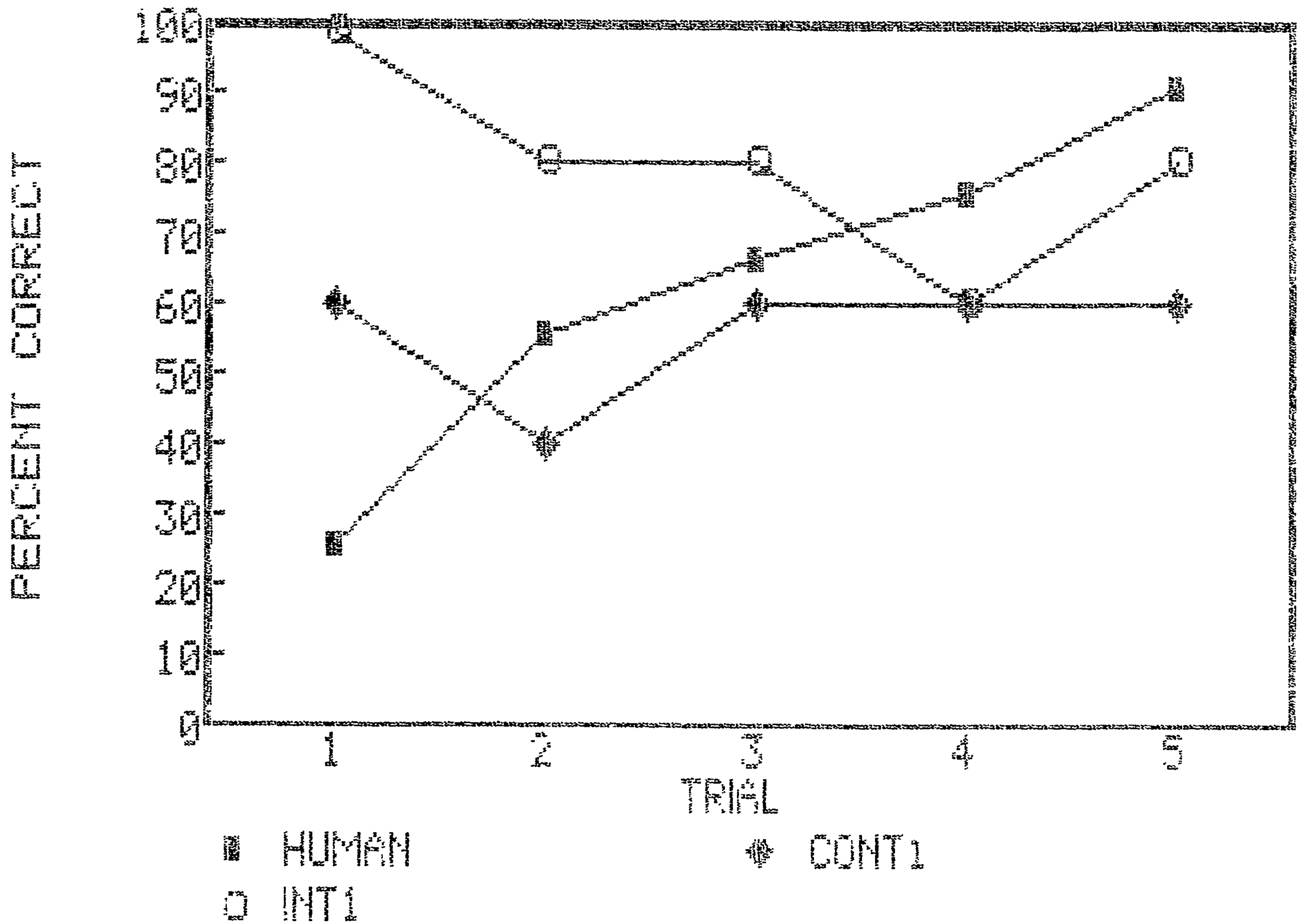

FIGURE 4 MEAN NOTE CORRECT RECOGNITION

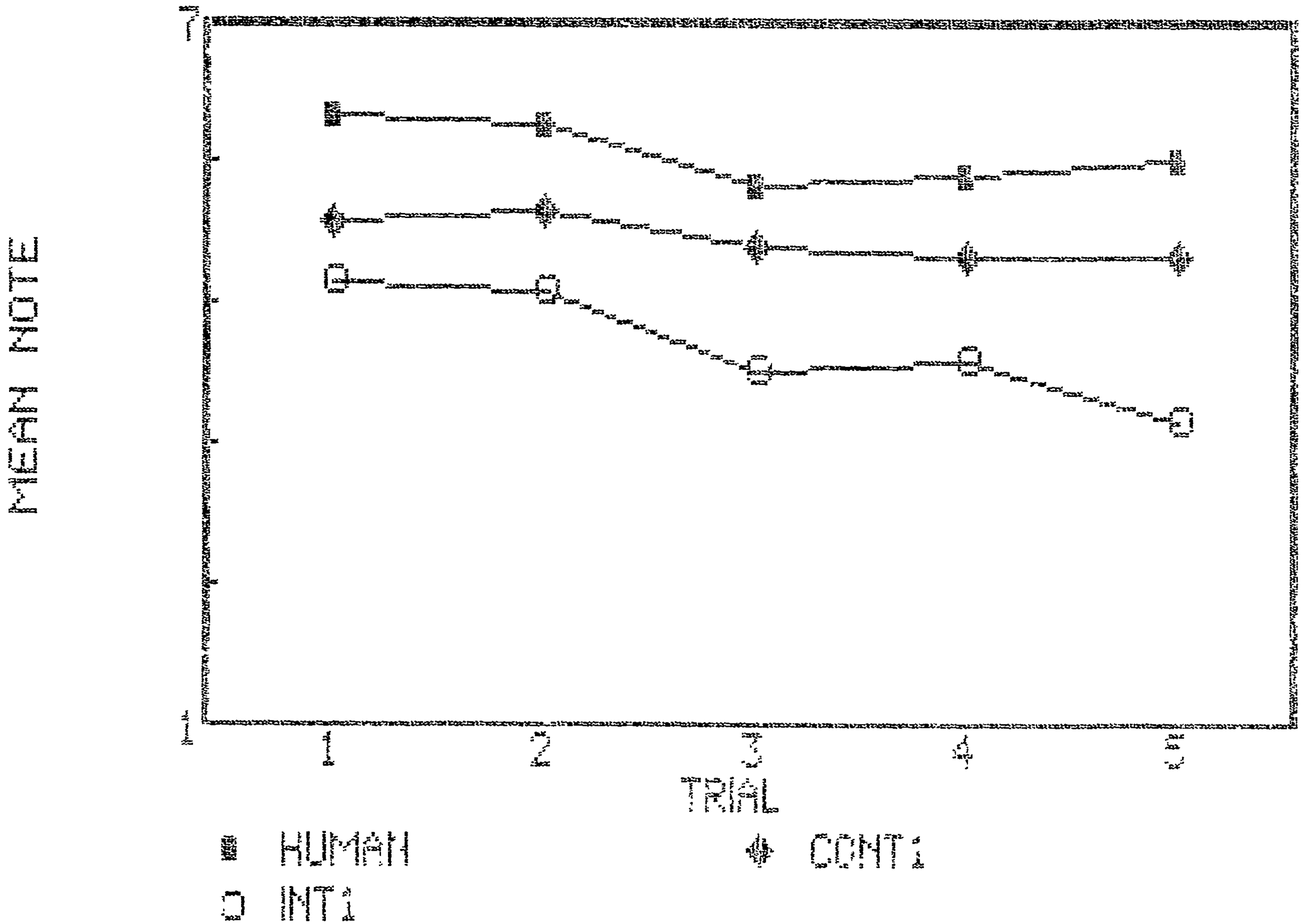

FIGURE 5 RELATIVE CONFUSION BY MELODY