Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyväskylä, Finland
Jukka Louhivuori, Tuomas Eerola, Suvi Saarikallio, Tommi Himberg, Päivi-Sisko Eerola (Editors)

Tune Recognition from Melody, Rhythm and Harmony

Tuire Kuusi

Department of Composition and Music Theory,

Sibelius Academy, Finland
tuire.kuusi@siba.fi

ABSTRACT

Earlier studies have shown that listeners recognise familiar tunes. With tunes that are manipulated in some way, melody has been shown to be more important for recognition than rhythm. The present study examined the importance of melody, rhythm and harmony for tune recognition by listeners with varying musical expertise.

Forty-six participants, divided into three groups according to their musical expertise, heard the first two phrases of familiar tunes in four different versions: *melody*, *rhythm*, *melody with harmony* and *rhythm with harmony*. The participants were asked to identify the tunes. A two-factor ANOVA was conducted with the four versions of presentation and the three groups of participants as experimental variables.

The study showed that both the versions of the tunes and the expertise of the participants were statistically significant factors for tune recognition; the professionals being best in tune recognition, and the amateur musicians being better than non-musicians. The rhythmic versions especially were recognised by the professionals, and, regardless of expertise, the melodic versions were easier to recognise than the rhythmic ones. Generally, harmony was not found to help tune recognition.

As a whole, the inexperienced listeners seemed to encode isolated details but were unable to process them, while the professionals were able to process information in a structural context. The differences in recognition of the rhythmic versions seemed to reflect the ability of professionals to extract temporal patterns and store temporal information.

I. INTRODUCTION AND AIMS

Both musically trained and untrained listeners can recognise familiar tunes. According to Serafine, Crowder & Repp (1984), the presence of text makes recognition easier for listeners. Yet tune recognition has been shown to be possible even when tunes are played without text (Hébert & Peretz, 1997; Prickett, 2000; Vongpaisal, Trehub & Schellenberg, 2006), regardless of the instrumentation, tempo (fast or slow) or lack of dynamic change (Warren, Gardner, Brubacker & Bashford, 1991). When tunes are played without text, recognition is based on both melody (pitch information) and rhythm (temporal information). According to Perezt and Morais (1989), melody and rhythm interact, even though they might be autonomously processed prior to integration. The presence of and attention to both is advantageous for recognition of tunes.

Tunes may also be recognised if they have been manipulated in some way or if the pitch information and temporal information are presented independently. Pitch information seems to be more important for recognition than temporal information (White, 1960; Deutsch, 1972; Moore & Rosen, 1979; Hébert & Peretz, 1997). Contour and scale step (or tone chroma) seem to be important as well, but when these two factors compete, the contour seems to be more important than the tone chroma (White, 1960; Deutsch, 1972; Dowling & Fujitani, 1971; Dowling & Hollombe, 1977; Idson & Massaro, 1978). Contour information can also be used if intervals are compressed or expanded, yet it is not easy for listeners to recognise tunes from contour information only; the exact interval sizes are important as well (Moore & Rosen, 1979).

Temporal information can be modified in two different ways: either by taking the shortest notes as a starting point and splitting the longer ones into as many repeated shorter ones as needed (e.g. Moore & Rosen, 1979) or by defining the same duration for each pitch (e.g. White, 1960). The former is easier for listeners, since it preserves the basic temporal structure and the original congruence between pitches and time signature, while the latter, more difficult modification does not. With the latter modification the melodic patterns may give misleading cues about metric information (see, e.g. London, 2004).

Studies of tune recognition have shown that excerpts of tunes can be relatively short and yet recognisable. In the study by White (1960), the tunes were recognised from the first six notes equally well as from the first 24 notes. In the study by Dalla Bella, Peretz and Aronoff (2003) the participants, when listening to familiar tunes, needed 7 notes at most to be confident of the familiarity of the tune. Yet extremely short excerpts are most likely recognised from timbral cues (Schellenberg, Iverson & McKinnon, 1999).

It has also been found that older participants are as good as younger ones in responding to similarities and differences in contour, rhythm or mode of melodies (Halpern, Bartlett & Dowling, 1998) and that increased musical expertise leads to better performance (Dalla Bella, Peretz & Aronoff, 2003), especially in conditions where the rhythm has been manipulated (Andrews, Dowling, Bartlett & Halpern, 1998).

The goal of the present study was to examine the importance of melody, rhythm and harmony for tune recognition in listeners with varying musical expertise. Specifically, the study examined how systematically the listeners recognised a tune from its first two phrases when an excerpt was played in different versions. A further goal was to determine how the recognition of different versions was affected by musical expertise.

In some earlier studies (White, 1960, Moore & Rosen, 1979, Schellenberg, Iverson & McKinnon, 1999) the participants were first given a list of the tunes to be used in the experiment. According to Hébert and Peretz (1997), this promoted top-down processing (that is, processing from memory to perception), since the list of tunes limited the number of possible answers. In the present study the participants were given no hints about which tunes were to be used (except that they were familiar). Hence, the study promoted bottom-up processing, that is, processing driven by sensory information.

II. METHODS

A. Participants

Forty-six individuals participated. They were divided into three groups according to their musical backgrounds. In the first group, labelled *professionals* (N = 14; 7 male), the participants were professional musicians who had experience with choral music (as choir conductors or professional musicians or music teachers who had also been singing in a choir for a long time). The average age was 38.9 years (range 22–62), and they had studied music professionally for 9.2 years (range 4–20). In the second group, labelled *amateurs* (N = 19; 5 male), the participants were choral singers with no professional musical training. The average age was 43.2 years (range 17–62), and they had studied music non-professionally for 11.9 years (range 0–45). The participants in the third group, labelled *non-musicians* (N = 13, 4 male), had not studied music (with the exception of music lessons at primary school), nor did any of them have music as a hobby (with the exception of listening to music). Their average age was 47.3 years (range 22–58).

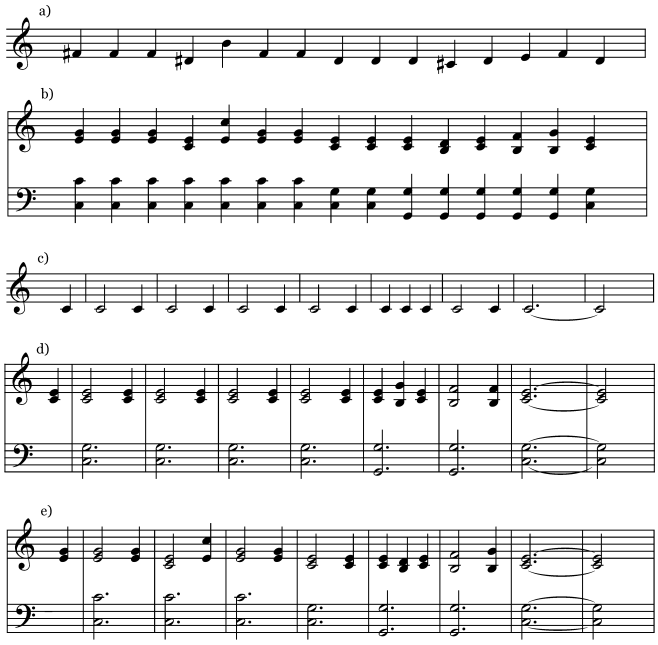
B. The tunes

The tunes used in the experiment had to meet two criteria. First, they had to be very familiar. Second, the rhythm of the tunes had to consist either of characteristic patterns or it had to have variation; tunes with a monotonous rhythm could not be accepted. Altogether 19 tunes were used. They can be arranged in eight groups (the number of songs in each group is given in parentheses): the national anthem (1), Christmas carols (5), folk songs (6), choral songs (3), commercial tunes (2), hymns (2) and other (3). The total number of songs in the groups is 22, since one folk song and one hymn were also choral songs and one hymn was also a Christmas carol.

The first two phrases were taken from each tune, meaning that each excerpt consisted of 12 to 23 notes (the average was 16.74 notes). The number of notes exceeded the critical number that has been reported earlier as being sufficient for tune recognition (see Introduction).

The tunes were played in four different versions: *melody*, *rhythm*, *melody with harmony* and *rhythm with harmony*. In addition, the tunes were played *correctly* (that is, with all parameters simultaneously) to determine whether the participants really knew the tunes used in the experiment. All versions were played using a digital grand piano sound (Model 1923 Steinway D from the PMI Old Lady sample library), without dynamic changes. The version *melody* indicated that all notes had the same duration (500 ms) and equal loudness. No hints of time signature or rhythm were given. *Melody with harmony* indicated that the tune was placed in the highest voice in a four-part arrangement with correct harmony, and all chords had equal length (500 ms). Since all notes in these versions were of equal length, the length of each excerpt varied according to the number of notes, and each lasted between 6.00 and 11.50 seconds.

The version *rhythm* indicated that the correct rhythm of the tune was played on a single pitch. A click (produced with the timbre of a metallophone) was added to mark the beginning of each measure. *Rhythm with harmony* was a four-part version of the tune with rhythm and harmony, but without the melody; again a metallophone click indicated the beginning of each measure. The rhythmic versions were played using a tempo that was close to the tempo in which the tune was usually heard or sung. The *correct* version included all parameters (rhythm, melody and harmony). The length of the versions *rhythm*, *rhythm with harmony* and *correct* varied between 4.05 and 12.31 seconds (average 7.09 seconds). Example 1 shows the four experimental versions (a-d) and the correct version (e) of "Lähteellä" (a folk song and a choir song).



EXAMPLE 1. The five versions of the tune "Lähteellä"; (a) *melody*, (b) *melody with harmony*, (c) *rhythm*, (d) *rhythm with harmony*, (e) *correct*.

The versions *melody* and *melody with harmony* retained the contour of the melody, while *rhythm* had no contour at all. The version *rhythm with harmony* contained some melodic information, since the highest pitch changed according to the chord. Yet the melodic elements differed from those included in the correct melody of the tune; they might also have reminded the listener of some other tune. In addition they provided a wrong contour. The transpositions of the melodies varied; however, it has been shown that listeners recognise familiar tunes regardless of the transposition (see e.g. Dowling, 1986).

C. The procedure

Each tune was presented in two different experimental versions, one rhythmic and another melodic, followed by the correct presentation. The correct presentations were played only after both experimental versions had been presented. Since it was possible that a previous version of a tune (even if not recognised) might have facilitated recognition of a later version (Hébert & Peretz, 1997), the order of the versions was controlled. The experiment consisted of 48 items. Of these, 42, consisting of 14 tunes, each played in three ways – two experimental versions and correctly – were analysed. Seven tunes were presented first as a rhythmic version (either rhythm or rhythm with harmony); another seven tunes were presented first as a melodic version (melody or melody with harmony). In addition the participants were played six 'false'

items: different versions of some other tunes, but the responses to these were not analysed.

Each item was played only once. After hearing the item, the participants were asked to respond either with the name of the tune, some words of the text (not necessarily from the beginning of the tune) or some description of the tune if they could not recall the exact name or text. Hence, the participants were asked to respond with expanded naming judgements.

When the responses were scored, the correct name or the correct words of the text were scored as 1; a correct description (e.g. Christmas carol, folk song) was scored as 0.5 and other responses were scored as 0. Some tunes were not recognised by the participants even when played correctly. All responses to an unrecognised tune were omitted from further analyses. The scores were summarised for each tune and each version and separately for each group of participants. Since the number of participants in the three groups was not constant and the number of tunes recognised by participants was not constant, the sums will be given as percentages: 75% indicates that a particular version of a tune was recognised by 75% of the participants who recognised the tune when played correctly.

When the results were analysed, a two-factor ANOVA was conducted with the four versions of presentation and the three groups of participants as experimental variables.

III. RESULTS

Figure 1 shows the percentages of recognised tunes in different versions for the three groups of participants. Generally the figure shows that the tunes were recognised when played correctly; more than 85% of the participants in each group recognised all tunes (professionals, 95.9%; amateurs, 93.6%; non-musicians, 86.5%). The figure also shows that the professional musicians recognised more tunes from all different experimental versions than did the other participants. The differences between the groups are small when the tunes were played correctly and larger with the four experimental versions. As can be seen, the tunes in the versions *rhythm* and *rhythm with harmony* were more difficult to recognise than were their versions as *melody* or *melody with harmony*. The ANOVA analysis confirmed the results: both the mode of presentation and the expertise of participants were statistically significant factors in explaining the responses ($F(3,83) = 28.493$ $p < .001$ and $F(2, 83) = 14.217$, $p < .001$ respectively).

Figure 1 also shows that professionals were much better in recognising the tunes from temporal information than were the less-trained or non-trained participants (the percentages for *rhythm* and *rhythm with harmony* were 50.5% and 37.6% among the professionals; 15.6% and 8.5% among the amateurs; 2.5% and 18.6% among the non-musicians). When comparing the two rhythmic versions, it can be seen that harmony did not help the musically trained participants to recognise the tunes, but helped the non-trained participants to some extent.

The differences between the groups were smaller with the melodic versions than with the rhythmic versions (the percentages for *melody* and *melody with harmony* were 76.1% and 81.3% among professionals; 67.4% and 68.0% among amateurs; 53.3% and 53.9% among non-musicians). Generally it seemed that harmony combined with the melody did not help the participants to recognise the tunes.

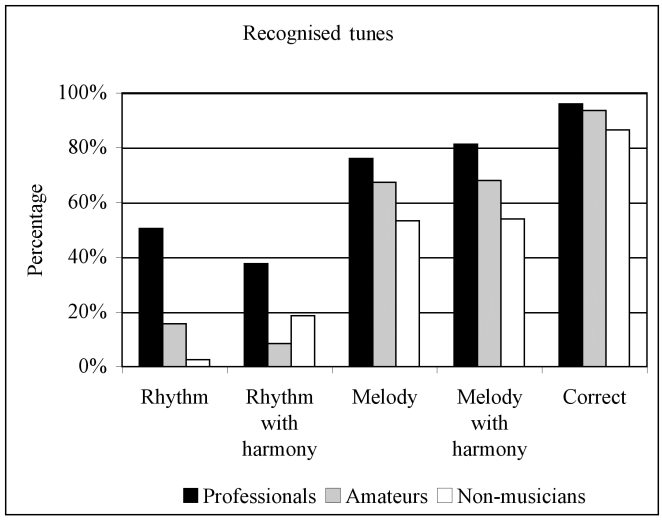


Figure 1. Percentages of the recognised versions of the tunes

In seven tunes a rhythmic version was presented before a melodic version; in another seven tunes the order was the opposite. The order was controlled to avoid any systematic error caused by the earlier presentation of a version of a tune (even unrecognised). The responses to the rhythmic versions and those of the melodic versions were analysed with regard to the order of presentation. The order did not seem to affect the responses; in other words, the earlier presentation of a version of a tune did not help the listeners to recognise another version of the tune (see Figure 2). The small differences in percentages are most likely owing to the tunes themselves, not to the mode of presentation.

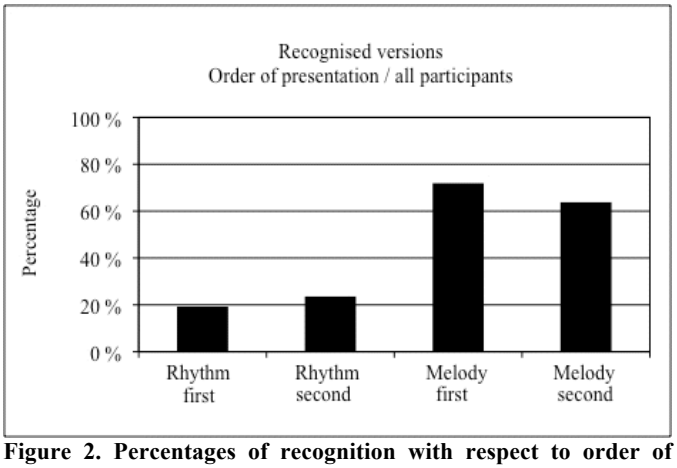


Figure 2. Percentages of recognition with respect to order of presentation.

IV. CONCLUSIONS

The present study showed that tune recognition was easiest when the tunes were played correctly, that is with correct pitches, rhythm and harmony. All participants, regardless of their musical background, recognised the national anthem. Yet all items in this study were not recognised; for example, the two commercial tunes were difficult for non-musicians

since, the percentage of their correct identifications was 80.75. One possible reason for this was that all the items were played using the piano timbre, and especially for these two tunes, the piano timbre did not fit. According to the encoding-specificity principle (Tulving & Thompson, 1973), the memory doe of a stimulus is better if cues at the time of retrieval match those at the time of encoding. It seems that transposition did not have as strong an effect on recognition as timbre; even though some commercial tunes are always heard in the same transposition, other tunes can be played and sung in varying transpositions.

It was also found that melody (with or without harmony) guided familiar tune recognition more strongly than rhythm. This finding is in accordance with earlier studies showing that pitch information is more important for tune recognition than temporal information (White, 1960; Deutsch, 1972; Moore & Rosen, 1979; Hébert & Peretz, 1997).

It is likely that the processing of pitch information becomes easier if the listener can understand the pitches and the relationships between the pitches in some (tonal) context, form meaningful combinations (in chords, parts of a scale, etc.), and then remember the combinations instead of the individual pitches. According to information recoding, first introduced by Miller (1956), listeners can reduce the amount of information to be retained by chunking subsets of more than one item into a single memory code. The same is true with temporal information: if the "shorts" and "longs" can be understood as meaningful subdivisions (and patterns), a metrical structure can be formed and a meter established (see e.g. London, 2004). The ability to process temporal information or pitch information requires an ability to store the information, and the ability to store the information requires the ability to perform auditive structuring (Karma, 1985). It has also been shown that experience and training modify long-term memory parallel to working memory and sensory memory (Schröger, Tervaniemi & Huotilainen, 2004); long-term memory can also affect the processing of sensory memory (Münte, Nager, Beiss, Schroeder, & Altenmüller, 2003).

Unlike the other participants, the professional musicians in the study were able to recognise tunes from rhythmic versions. This difference from other participants might indicate differences in information processing. The participants who were highly trained musically were able to store temporal information, since they were able to extract rhythmic patterns from the series of sounds and then remember the rhythm and repeat it in their minds The metallophone click might have helped them to establish the groupings (time signatures) and thus establishing a meaningful meter. The musically less-trained or untrained participants, on the other hand, were not able to process the temporal information; hence, they forgot it and could not recognise the tune.

Even though harmony was assumed to help tune recognition (since it gave the tonal context), this was not the case; on the contrary, the presence of harmony seemed to distract from the recognition of tunes in rhythmic versions among the musically trained participants and did not particularly help in their recognition of the melodic versions either. With rhythmic versions this was most likely because the highest pitch of the version *rhythm with harmony* included melodic elements and provided misleading information. The musically trained participants obviously tried to use the pitch information (actually, melody) as a cue, but since the information was incorrect ("wrong melody"), it did not help; instead, it distracted in the processing of temporal information. The misleading pitch information did not, however, disturb the non-musicians; since they were not able to structure and store temporal information, the harmony added to the rhythm and the "wrong melody" did not distract them. Instead, the harmony added to the rhythm seemed to help to some extent, possibly because it provided additional (pitch) information.

As a whole, the study confirmed the results obtained in earlier studies: both musically trained and untrained listeners can recognise familiar tunes, and melody is more important for recognition than rhythm. The role of harmony seemed to be of limited importance, yet further study of the role of harmony would be of interest. The main problem with harmony is that it always includes pitch information forming melodic elements (which seemed to dominate); hence, the main question is whether listeners can concentrate on harmony without paying attention to the melody. In the present study (and in earlier studies as well) the tunes were played from the beginning. It would be interesting to study tune recognition with excerpts in typical melodic, rhythmic or harmonic patterns. In addition the role of listener expertise could be studied by examining how quickly listeners with varying music educational backgrounds recognise familiar tunes.

REFERENCES

Andrews, M.W., Dowling, W. J., Bartlett, J. C., & Halpern, A. R. (1998). Identification of speeded and slowed familiar melodies by younger, middle-aged, and older musicians and nonmusicians. *Psychology and Aging, 13*, 462–471.

Dalla Bella, S., Peretz, I., & Aronoff, N. (2003). Time course of melody recognition: A gating paradigm study. *Perception & Psychophysics, 65*, 1019–1028.

Deutsch, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics, 11*, 411–412.

Dowling, W. J. (1986). Context effects on melody recognition: Scale-step versus interval representations. *Music Perception, 3*, 281–296.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49*, 524–531.

Dowling, W. J., & Hollombe, A. W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception & Psychophysics, 21*, 60–64.

Halpern, A. R., Bartlett, J. C. & Dowling, W. J. (1998). Perception of mode, rhythm, and contour in unfamiliar melodies: Effects of age and experience. *Music Perception, 15*, 335–356.

Hébert, S., & Peretz, I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory & Cognition, 25*, 518–533.

Idson, W. L., & Massaro, D. W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics, 24*, 551-565.

Karma, K. (1985). Components of auditive structuring – Towards a theory of musical aptitude. *Council for Research in Music Education, Bulletin 82,* 1–13.

London, J. (2004). *Hearing in time. Psychological aspects of musical meter.* New York: Oxford University Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Moore, B. C. J., & Rosen, S. M. (1979). Tune recognition with reduced pitch and interval information. *Quarterly Journal of Experimental Psychology, 31*, 229–240.

Münte, T. F., Nager, W., Beiss, T., Schroeder, C., & Altenmüller, E. (2003). Specialization of the specialized: Electrophysiological investigations in

professional musicians. *The Neurosciences and Music. Annals of the New York Academy of Sciences, 999*, 131–139.

Peretz, I., & Morais, J. (1989). Music and modularity. *Contemporary Music Review, 4*, 279–293.

Prickett, C. A. (2000). A Comparison of the Basic Song Repertoire of Vocal/Choral and Instrumental Music Education Majors. *Journal of Research in Music Education, 48*, 5–9.

Serafine, M. L., Crowder, R. G., & Repp, B. H. (1984). Integration of melody and text in memory for songs. *Cognition, 16*, 285–303

Schellenberg, E. G., & Iverson, P., & McKinnon, M. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review, 6*, 641–646.

Schröger, E., Tervaniemi, M., & Huotilainen, M. (2004). Bottom-up and top-down flows of information within auditory memory: Electrophysiological evidence. In C. Kaernbach, E. Schröger & H. J. Müller (Eds.), *Psychophysics beyond Sensation: Laws and Invariants of Human Cognition* (pp. 389–407). Hillsdale, New Jersey: Erlbaum.

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.

Vongpaisal, T., Trehub, S. E., & Schellenberg, E. G. (2006). Song recognition by children and adolescents with cochlear implants. *Journal of Speech, Language, and Hearing Research, 49*, 1091–1103.

Warren, R. M., Gardner, D. A., Brubacker, B. S., & Bashford, J. A. Jr. (1991). Melodic and nonmelodic sequences of tones: Effects of duration on perception. *Music Perception, 8*, 277–290.

White, B. W. (1960). Recognition of distorted melodies. *The American journal of psychology, 73*, 100–107.